Mechanistic Simulation of Normal-Tissue Damage in Radiotherapy

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Eva Sofie Rutkowska

December 2010

Mechanistic Simulation of Normal-Tissue Damage in Radiotherapy

Eva Rutkowska

Radiobiological modelling is essential to the progress of clinical radiotherapy. The use of robust normal-tissue complication probability (NTCP) models has the potential to allow truly individualised treatment plans where the tumour control probability is not limited unnecessarily by the sometimes conservative application of techniques based on the response of a population.

In this thesis NTCP models and methods for analysing clinical data are tested on data simulated by a 3D mechanistic model of normal-tissue damage. This mechanistic model was developed to represent local tissue damage by functional subunit inactivation, and the overall organ response by a critical functioning volume (CFV). The size of the CFV varies between organs and depends on the volume effect, i.e. how large volumes of tissue damage are necessary to cause a complication. The model complexity was guided by the degree of information available about the pathogenesis of radiation-induced side-effects.

The model was used to generate pseudo-clinical datasets, which typically consisted of dose-volume histograms (DVHs) and binary complication information, for a large number of simulated treatments. Because all dataset characteristics and the 'biology' of the organs were known, studies on the simulated data could give insights into data analysis methodology.

It was demonstrated that correlation analyses between dose-volume parameters and outcome are strongly influenced by dataset characteristics, but that information about the volume effect of the organ can be gained if care is taken to identify all factors influencing the observed correlations.

The relative performance of several DVH-based NTCP models was explored for different levels of confounding factors, and it was found that model performance was influenced more by confounding factors than by the choice of model.

The 3D mechanistic model of normal-tissue damage, developed in this thesis, is a powerful tool for studying data analysis methodology and is also useful as a framework for summarising the radiobiological knowledgebase of normal-tissue effects in radiotherapy.

There is great potential to develop the model further to include e.g. non-local effects of irradiation, time effects, and several FSU populations, related to different endpoints, co-existing in one organ.

Acknowledgements

I would like to thank my supervisors, Dr Colin Baker and Prof Alan Nahum, for making my time as a PhD student so rewarding, by allowing space for me to try ideas and develop a deep interest in my project, while providing steady support and encouragement as well as using their experience and insight to challenge my understanding. I am truly grateful for the time they have invested to ensure the quality of my work and my progress as a researcher.

Material to this progress have also been the resources and support provided by the Graduate School and Dr Ian Smith at the Computing Services, as well as the financial support by the School of Health Sciences at the University of Liverpool.

The friendship and cooperation of my colleagues in the medical physics reseach group at Clatterbridge Centre for Oncology; Jothy Basu, Mekala Chandrasekaran, Vanessa Panettieri, Juan Pardo-Montero, Alison Scott, Julien Uzan and Aisyah Yusof Azuddin mean a lot to me. What would these years have been without our lengthy discussions and culinary delights?

I would also like to remember the valuable support offered by Dr Isabel Syndikus, Dr Giovanna Gagliardi and Dr John Fenwick.

Above all, I want to thank my husband, without whom I would not have ventured into this project, for all that he has given up for me to be able to achieve my goals, and for his love and friendship which mean the world to me. I am also immensely grateful to my parents who have loved and supported me without requiring success, and to my sisters and their families.

Contents

Chapter 1

Introduction

1.1 Pathogenesis of radiation side effects

Modern radiotherapy techniques generally cause no or mild side effects, although a few patients experience serious morbidity following radiotherapy [1]. Irradiation of the heart and spinal cord can lead to the lifethreatening condition ischaemic heart disease [2] and paralysis [3] respectively, and poor lung function [4], rectal injury [5], dysphagia [6, 7, 8] etc., caused by irradiation, can lead to severely reduced quality of life. However, treatment schedules have been developed which are associated with a low risk of such serious side effects [1]. This limits the radiation dose which can be delivered to the tumour since increasing the dose to the tumour generally leads to higher doses in the normal tissues around it. It is therefore important to study the tolerance of normal tissues to irradiation in order not to limit the dose unnecessarily in patients with a low risk of developing complications since this would decrease the chances of achieving local tumour control.

1.1.1 Tissue damage

Like chemotherapy, radiotherapy is a cytotoxic treatment where the goal is to prevent cell proliferation in the tumour by inducing reproductive cell death. Since also normal tissues are unavoidably irradiated to some extent, cell dysfunction and cell death can lead to tissue damage and organ failure. Radiation causes damage through ionisation, which transforms molecules into unstable free radicals. DNA molecules can be damaged directly, but 70% of the biological damage by a photon beam is caused by free radicals,

formed from water, which react with DNA and other biological molecules [9, p. 56].

Generally cell death following radiotherapy is in the form of mitotic cell death, i.e. the proliferative capacity of the cell is lost. However, all the irradiated cells uninterruptedly perform their normal function throughout their normal life span [9, 10]. The mitotic process is the most radiosensitive cell function, since the chromosomes in the cell nucleus, on whose integrity mitosis is dependent, are unique; in contrast cell plasma components, maintaining other cell functions, generally exist in large numbers and are easily regenerated by the cell [9, 11]. Therefore it is reasonable to assume normal cell function after radiotherapy until cell death, even in cells that have been irreversibly damaged [9, ch. 6]. Also, surviving cells can be assumed not to contribute to tissue damage [12]. Damaged stem cells, however, fail to proliferate through mitosis so that the tissue is eventually depleted of its cells. For the lung for example, this means a manifestation of the normal-tissue effect 2 to 4 months after radiotherapy [10, 13].

The viability of a cell after irradiation can be assessed in vitro by its colony-forming ability [9, ch. 6]. A cell which does not retain full proliferative capacity will not contribute to future generations of cells. The intrinsic radiosensitivity of a cell line is often evaluated based on dose-response curves, created by counting the fraction of cells which retain their colony-forming ability after delivering increasing doses of radiation to groups of such cells. Tumour cells do not seem to differ from similar normal cells in terms of intrinsic radiosensitivity; the mechanisms inducing apoptosis after damage to vital chromosomes might not function normally in tumour cells, but cell survival ultimately depends on the colony-forming ability, which is lost during irradiation even if damaged cells do not disappear immediately. However, the radiation response of a cell also depends on its environment, and in this respect there are differences between tumour cells and normal-tissue cells [9, ch. 12].

Not all organs and tissues are built of proliferating cells, but rather mature post-mitotic cells [13]. In some organs, though, such cells have the capacity to start proliferating if necessary, and in other organs a constant renewal of the cell population takes place, to compensate for natural cell loss. Tissues without any cell population turnover are especially radiation resistant since no proliferative capacity is necessary. However, the common conclusion that tissues with slow turnover are generally more radioresistant than tissues with fast turnover is misleading [9, p. 217]; the only difference

is that serious tissue damage appears earlier for the latter cell type, if irradiated with a high dose. However, since tissue integrity depends on the composite effect of cell loss, cell proliferation and cell migration, cell populations among early responding tissues with especially short cycle might experience more severe effects if cell loss is much faster than cell migration, as illustrated by Partridge [14].

Cell kill by radiation can be described mathematically with the linear-quadratic (LQ) model [11, 15], where the surviving fraction (SF) of cells, irradiated with dose d, is

$$SF = \mathrm{e}^{-\alpha d - \beta d^2}. \tag{1.1}$$

Here α and β are radiosensitivity parameters associated with dose-rate independent and dose-rate dependent cell-killing mechanisms respectively [16]. If several (n) equal fractions of dose d are delivered, separated by enough time to allow complete repair of sublethal damage, and at high dose-rate so that this repair is negligible during each treatment fraction,

$$SF = \mathrm{e}^{-\alpha n d - \beta n d^2} = \mathrm{e}^{-\alpha n d\left(1+\frac{d}{\alpha/\beta}\right)}. \tag{1.2}$$

The fraction α/β is related to the fractionation sensitivity of the tissue, and is around 1-5 Gy for late responding tissues and around 8-12 Gy for early responding tissues [17]. Tumour cells generally also have a high α/β.

Since the total effect of a fractionated treatment depends on the fraction size, doses are often normalised to the equivalent total dose in 2-Gy fractions which, according to the LQ model, would give the same cell kill:

$$\mathrm{EQD2} = nd\,\frac{1+\frac{d}{\alpha/\beta}}{1+\frac{2}{\alpha/\beta}} \tag{1.3}$$

After normalisation, doses given in fractions of different sizes can be compared in terms of EQD2.

This traditional framework of radiation-induced tissue damage, called the target cell theory [18, 19], has been criticised for being too simple [20, 21]. It is now clear that, even in so-called late responding tissues, radiation triggers processes which develop throughout the latent period before the manifestation of clinical symptoms. The cellular homeostasis is perturbed and cytokines are released, sending messages between cells,

the effect of which is highly unpredictable. Cytokines are associated with the induction of free radicals and inflammation which can cause additional tissue damage, or indeed protection. However, despite the recent interest in these mechanisms it is recognised that, although direct radiation cell kill is not the only cause of radiation pathology, it is probably the most important [22, 20, 21].

The importance of free radicals for radiation-induced tissue damage was mentioned above. Interestingly several other pathologies are attributed to the tissue damage caused by free radicals [23, p. 20]. Also, the normal-tissue reactions to radiation are not unique but are observed as a result of attack by various agents [24, 21]. An important difference, though, might be the repetitive nature of the injury caused by fractionated radiotherapy, which is delivered under continually changing pathological conditions [25]. Considering the biological events caused by irradiation it is natural to expect the normal-tissue response to radiotherapy to be influenced by pre-existing co-morbidity as well as by other treatment modalities such as surgery and chemotherapy [25, 13]. The clinical expression of radiation toxicity also seems to be influenced by post-treatment pharmacological interventions and diseases [25].

1.1.2 Organ failure

It is important to differentiate between tissue damage and morbidity [21]. In some organs a complication occurs if any tissue damage at all is observed. This is the case for moist desquamation in the skin and ulceration in the rectum. This kind of complication has been named anatomical/structural organ failure [26]. In other organs, and for other endpoints, the link between tissue damage and organ failure is less straightforward. In some organs the function is performed "in parallel" by multiple subunits, such as the alveoli in the lung, nephrons in the kidney and acini in the liver. This means that with increasing tissue damage there will be a gradual loss of organ function since the undamaged parts of the organ are unaffected but must cope with an increasing workload [19]. Moreover, these organs often have a functional reserve and a complication does not occur until a significant fraction of the organ has been damaged [27].

Marks [27] also discusses the relative importance of different tissues within a given organ. Organ failure can result from excessive damage to any essential cell population within the organ, and perhaps even outside the organ. Thus, the complication probability

can have a complex dependence on the dose distribution. Organ dysfunction might occur as a result of a large fraction of radiosensitive cells being damaged through the irradiation of large volumes to relatively low doses, but if instead the irradiated volume were limited, but small regions received very high doses, a cell population which is less radiosensitive might be affected. Also, since similar symptoms can arise from damage to different cells it is often difficult to identify the pathogenesis of a given complication [26].

In many organs different regions contribute heterogeneously to organ function, and even more so if diseased. As a consequence, irradiating different parts of an organ could have different effects. The bladder, femur, lung and kidney are examples of organs commonly exhibiting complications from radiotherapy which have internal structures which might lead to a heterogeneous response to irradiation [27].

On a population basis the prescription dose is guided by the tolerance of the organs at risk [28]. The rationale of conformal radiotherapy is that the tolerance dose of an organ (assuming homogeneous partial organ irradiation) often increases if the irradiated volume decreases. This phenomenon is called the volume effect and can be caused by several different mechanisms [1, 29]. Firstly, the dose-volume response of an organ strongly depends on its functional reserve. If the functional reserve is large the complication probability is associated with the overall tissue damage, whilst if it is small even small volumes of tissue damage cause morbidity. Alber and Nüsslin called this the volume effect of the first kind [30]. Secondly, for anatomical/structural damage, tissue repair often relies on migration of viable cells into the damaged volume, and in this case there is a saturation of damage when distances within the damaged volume exceed the distance over which a cell can migrate. It is therefore not surprising that the critical irradiated volume seems to be independent of the size of the individual [26]. Thirdly, the stochastic nature of radiation damage can cause a volume effect in some tissues [1, 29] (called the volume effect of the second kind by Alber and Nüsslin [30]). This would be the case if a complication were triggered by a certain amount of contiguous tissue damage, since, if a larger volume is irradiated the likelihood increases that, somewhere in the organ, a critical contiguous volume is damaged. Finally, it is also possible that the radiosensitivity of a tissue could increase with increasing irradiated volume [29], perhaps due to secondary injury caused by oxygen radicals produced by inflammatory cytokines (which can also cause 'out of field' effects) [21]. If more cells are irradiated the

amount of cytokines in the irradiated volume could increase, leading to the additional tissue damage, and also a certain level of damage outside the irradiated volume [21]. More than one of the effects listed above could be important for the response of a single organ. It is clear that not only cell kill, but also the architecture and physiology of the organ at risk, are important for the risk of complications from radiotherapy [26, 21].

As an example, reducing the volume irradiated for an organ with a small volume effect can sometimes reduce the complication probability considerably, as pointed out by Fenwick and Nahum [31], however, such a reduction in irradiated volume would not allow a large increase in maximum dose. This is because the benefit of reducing the treated volume comes from the reduced complication probability due to stochastic effects (third mechanism mentioned above), whilst, since the volume effect is small, the complication probability would increase sharply with the maximum dose (first mechanism). Generally, however, an organ with a large volume effect would be expected to benefit more from a reduction in treated volume, compared to an organ with a small volume effect, and such a reduction in treated volume would allow an increase in dose.

1.1.3 Functional subunits

A functional subunit (FSU) is the tissue volume which can be regenerated by a single stem cell. Consequently, if all FSUs retain at least one stem cell no tissue damage occurs from irradiation. According to Withers *et al.* [29] "organ function depends upon the aggregation of cells into functional subunits". Each FSU performs the same function, and if damaged cannot be rescued by stem cells from neighbouring FSUs.

In addition, the architecture of some tissues can be described in terms of structural subunits which perform the tissue function independently (but might need several surviving stem cells). Importantly, the FSU is defined for tissue repair purposes, whilst the structural subunit is anatomically defined. In some cases the structural subunit might be identical to the FSU, as seems to be the case for the nephron in the kidney; on the other hand, one tubule of the testis (the structural subunit) consists of multiple FSUs; and in some cases the FSU includes several structural subunits (e.g. intestinal crypts, hair follicles) [29].

The relationship between tissue damage and organ dysfunction depends on the organisation of the FSUs in the tissue [21]. It is sometimes helpful to compare the FSU

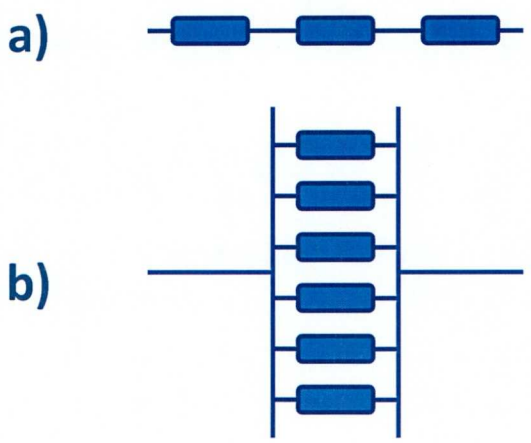


Figure 1.1: Schematic illustration of a) serial and b) parallel organisation of FSUs.

to an electrical component which can be connected to other components in series or in parallel; see figure 1.1. A serial organ architecture is characterised by FSUs organised in line, and the organ function is often structural which means that the loss of any one of the FSUs will cause morbidity. In a parallel organ, on the other hand, the FSUs contribute to organ function independently and the loss of a few FSUs does not lead to total loss of organ function. Consequently, serial organs, like the spinal cord, have a very small functional reserve, whilst parallel organs, like the lung, can have a large functional reserve.

These different types of tissue architecture have also been compared to a chain and a rope, respectively, since the rope has a certain redundancy of fibers whilst the chain depends on all links being intact [32]. However, since the structure and function of different organs differ greatly, such analogies can only be taken so far and though the concept of the FSU seems flexible enough to be useful in the description of most organs, the analogy with electric circuits has severe limitations. It is probably more useful, and less misleading, to characterise an organ in terms of the size of its functional reserve, than as a serial/parallel organ [33].

Since by definition an FSU can be regenerated by a single surviving stem cell the dose resulting in tissue damage depends not only on the intrinsic radiosensitivity of the cells, but also on the number of stem cells in an FSU [34, 21]. If the FSUs have few

stem cells they are more likely to loose them all if irradiated, and thus it is more likely that some FSUs will be destroyed. In contrast, if there are many stem cells in each FSU all FSUs might survive, in which case the tissue retains its integrity even if a lot of cells are killed.

As described in section 1.1.2 the function of an organ generally depends on more than one cell population, and though one kind of FSU might include different kinds of cells, e.g. type I and type II pneumocytes in the pulmonary epithelium, in some cases it is more appropriate to describe the tissues of an organ in terms of different FSU populations. Such FSU-pupulations can have different architectures and functional reserves; a typical example is the parenchyma and the vasculature [35], or similarly the parenchyma and the bronchial tree in the lung. If all pneumocytes of some alveoli are killed by radiation the effect is a corresponding loss of parenchymal FSUs which largely function independently of each other. In contrast, the structure of the bronchial tree makes damage to large bronchi in the central lung much more detrimental than damage to the peripheral bronchi, since all airways downstream will be affected [36, 37]. Evidence from radiation-induced rectal injury indicates that such populations of FSUs can be interdependent, and damage to the microvascular endothelium cause secondary damage to the epithelial cells of the mucosa [38].

1.2 Treatment plan evaluation

Evidence-based medicine requires modelling on experimental and clinical data to guide treatment planning, for e.g. drug or radiation prescription. Typically the data is binary: success versus failure of the treatment, or presence versus absence of a treatment-induced side-effect. There are standard statistical methods to model binary data, in terms of the probability of causing a certain effect, given the value of a continuous variable such as dose; see section 3.1.1. These methods are generally directly applicable when modelling tumour control probability (TCP), since each patient can be characterised by success/failure of the treatment for a given radiation dose prescribed to the tumour. Under these conditions the TCP is an s-shaped function of the prescription dose, which can be parameterised e.g. by its slope and location on the dose axis.

In radiotherapy the treatment plan is also evaluated in terms of its normal-tissue complication probability (NTCP). If a whole organ, or a part of an organ, is irradiated to a uniform dose the NTCP can be modelled as an s-shaped function of the dose to the organ, similarly to the TCP case. However, it is desirable to keep the dose to normal tissues in the vicinity of a radiotherapy target as low as possible, without compromising tumour control, which generally causes the three-dimensional (3D) dose distribution in the normal tissues to be highly heterogeneous. In this case the binary outcome (normal-tissue complication) depends, not on a single continuous dose variable, but on a 3D distribution of dose.

Thus, a large amount of data must be considered when modelling NTCP [35], and if standard statistical methods of modelling binary data are to be used, this problem needs to be simplified. Conventionally, the 3D dose distribution in an organ has been reduced to a dose-volume histogram (DVH), which summarises the dose distribution in terms of the volumes receiving a range of doses. Any spatial information in the treatment plan is lost in this step.

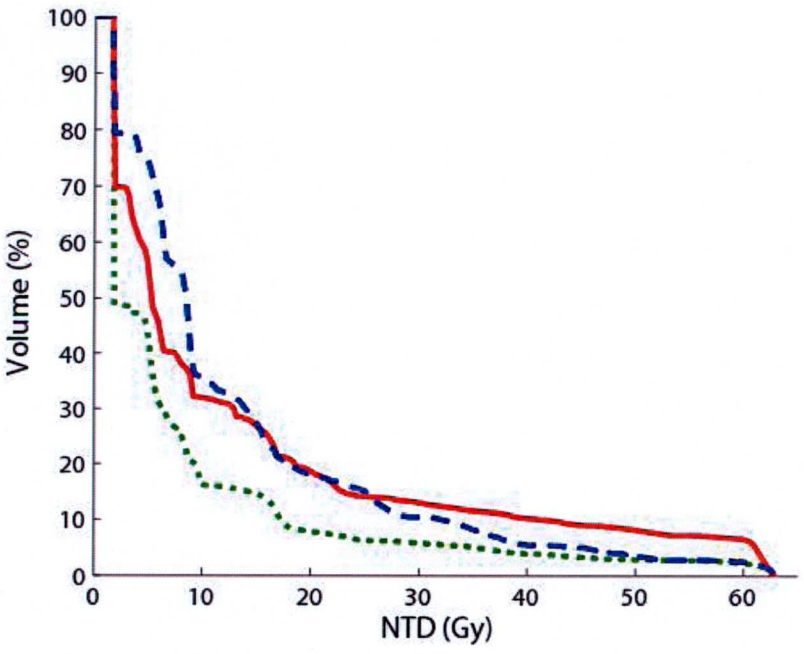


Figure 1.2: Three cumulative DVHs for an organ at risk. The dotted DVH is clearly the safest alternative since all dose levels are given to smaller volumes, but since the full and the dashed lines cross over a visual evaluation of the DVHs is not enough to rank one as safer than the other.

For common organs at risk (OAR) cumulative DVHs are useful as graphical evaluations of treatment plans, and unless the DVHs corresponding to two treatment plans cross over, the plan resulting in lower volumes treated to lower doses should be associated with a lower NTCP; see figure 1.2. However, DVHs are still too complex for standard statistical modelling, and are often summarised by single DVH metrics such as the mean dose, maximum dose, V_x (the volume receiving at least x Gy) and D_x (the lowest dose to the hottest $x\%$ of the OAR). Other metrics often used to summarise the DVH are an effective volume (V_{eff}) and an equivalent uniform dose (EUD), which, according to some function, reduce the DVH to a single value but are derived from the full DVH; see section 2.1.1.

1.3 Aims

As discussed in section 2.2, the risk of causing radiation-induced side-effects can be quantified based on direct statistical analyses of collected clinical and dosimetric data, or by the application of a conceptual risk model to the treatment plan. The former approach involves limitations of generality since it is based on observed outcomes in a limited clinical context, and for the second approach to be reliable all radiobiological assumptions in the model must be valid.

One way of evaluating the generality of conclusions from a clinical study or the validity of a conceptual model is to perform animal studies where different confounding factors can be controlled and untried dose distributions can be applied to normal tissues without concern for the risk of complications. Examples of the contribution of animal models to current understanding of organ response to irradiation are the studies of the spinal cord volume effect, and the studies of the spatial variability of tissue response in the lung, performed on mice and rats [39, 40, 41].

This thesis introduces an alternative method of exploring the interface between theoretical radiobiology and clinical radiotherapy. A mechanistic 3D computer model of an organ/normal tissue is developed, with local tissue damage represented by FSU inactivation. The volume effect (determining organ injury) is represented by a variable "critical functioning (fractional) volume" (CFV). Unlike DVH-based NTCP models, this model can include spatial effects of dose distributions, and rather than to evaluate treat-

ment plans its purpose is to simulate radiotherapy treatments, generating pseudo-clinical complication data for a patient depending on the choice of a set of parameters. Large numbers of treatments can be simulated in order to test data analysis methodologies. The model can also be a tool for investigating the behaviour of existing NTCP models for arbitrary dose distributions.

The model is described in detail in chapter 4, as well as the methods of choosing parameter values and generating dose distributions. In chapter 5 clinical datasets are simulated for hypothetical organs in order to explore the influence of dataset characteristics on correlations between dose-volume parameters and outcome.

Further, in chapter 6 the parameter values of the 3D model are adapted to simulate two real organs, lung and rectum, as closely as possible, and the radiobiological knowledge base necessary for mechanistic modelling is explored. After having tuned the model parameters to these organs, in chapter 7 clinical datasets are again simulated in order to compare the performance of different DVH-based NTCP models for different levels on inter-patient variation in radiosensitivity and health status.

The 3D model is developed as a biological mechanistic model, aiming to represent the mechanisms of normal-tissue damage. The level of complexity has been guided by the degree of information available about the pathogenesis of radiation-induced side-effects, by a principle of generality which allows it to be applied to different organs, and at the same time a desire to accommodate organ specific features of anatomy and physiology which influence the organ response to irradiation [26]. One aim of this work is to gain further understanding of the factors influencing the risk of radiation-induced morbidity, by adjusting the parameter values to experimental and clinical findings and comparing the resulting simulations with clinical studies. Thus, like other attempts to model normal-tissue effects mechanistically, this work aims to give a glimpse into the black box of radiotherapy.

The nature of the model output, as a binary response, makes the model especially suitable for simulating clinical data rather than for treatment plan evaluation (although treatment plan evaluation would be possible through averaging over patient variability). Therefore most of this work focuses on generating representative clinical datasets which are not limited in terms of numbers, treatment technique or even complication rate, in contrast to real clinical data. Since the 3D model is unlikely to be a perfect representation

of the normal-tissue response, such results do not so much predict the normal-tissue response under such unexplored conditions, as provide a tool for testing data analysis methodology. Little is known about how the measures currently used in treatment plan evaluation relate to the biology of the individual being treated or the population which these measures are taken from. By applying data analysis methods to datasets where the 'biology' is known the link to population statistics is explored.

Chapter 2

Normal-Tissue Complication Probability Models

2.1 DVH-based NTCP models

DVH metrics are often used as constraints in treatment planning and they can be linked to NTCP by fitting s-shaped functions to clinical data. There is a growing interest in the clinical application of such models, especially in combination with a DVH reduction scheme which takes the doses to all volumes into account. Most DVH-based NTCP models which are suitable for clinical use reduce the DVH to a single summary measure under some assumptions of an organ-specific volume effect, reasoning that the architecture of an organ can make it especially vulnerable to specific features of a dose distribution. Moreover, DVH reduction schemes implicitly or explicitly assume a local dose effect mechanism. For example, if a tissue were believed to be destroyed above a threshold dose, x Gy, but spared for doses below this threshold, a step function would be adopted for the DVH reduction and the DVH metric of importance would be V_x (the volume receiving at least x Gy). Similarly, if the mean organ dose is chosen as summary measure, the effect on the tissue is assumed to increase linearly with increasing dose. Alber [42] showed that treatment plan optimisation can be based on minimising the summary measure of choice, without having to calculate NTCP. However, it is often useful to link the summary measure to an actual risk estimate.

Generally one or two parameters are used in the DVH reduction expression. The resulting summary measure is then related to the NTCP with an s-shaped function

which uses two additional parameters, e.g. the value of the summary measure resulting in 50% complication probability and the slope of the curve at this point [43, 44]. The most common s-shaped functions for modelling binary data are the logit and the probit (cumulative normal) functions [45, ch. 3], whilst in NTCP modelling also an exponential binomial or Poisson function is often used [46, 20, 47], since this provides a direct link to radiation cell kill through the LQ model [15]. As reviewed by Bentzen and Tucker, it has been suggested that the exponential function is a superior choice due to its mechanistic aspect [48], whilst the advantage of the logit function is its computational simplicity [49, 45]. Moreover, the latter provides a natural method of incorporating non-dosimetric predictors; see section 7.2.3. However, the most commonly used NTCP model, the Lyman-Kutcher-Burman (LKB) model [50, 51] uses the probit function. The probit function has a connection to radiobiology since it approximates the cumulative binomial function which has been used to model the probability of damaging an FSU [52]; see section 2.2.

For a chosen summary measure, ϕ, the probit function gives the NTCP:

$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x/2} \mathrm{d}x, \tag{2.1}$$

where

$$t = \frac{\phi - \phi_{50}}{m \cdot \phi_{50}}. \tag{2.2}$$

Here, the value of ϕ associated with 50% NTCP, ϕ_{50}, and the relative standard deviation, m, are two parameters which are fitted to clinical data. As pointed out by Tsougos *et al.* [53], this model assumes a normally distributed probability distribution for ϕ_{50}, and the slope of the curve (and thus m) depends on its variance. The s-shaped curve given by this function is illustrated in figure 2.1.

Some of the local dose effect measures most commonly used in equation 2.2 are summarised below (section 2.1.1 and 2.1.2). It is the assumptions about the volume effect and local dose effect mechanisms which differ between the models. A commonly used model with a 'reversed' structure is the relative seriality model (section 2.1.3). With this model doses are first translated to probabilities, and then a probability-volume histogram is reduced to an NTCP, based on assumptions about the volume effect.

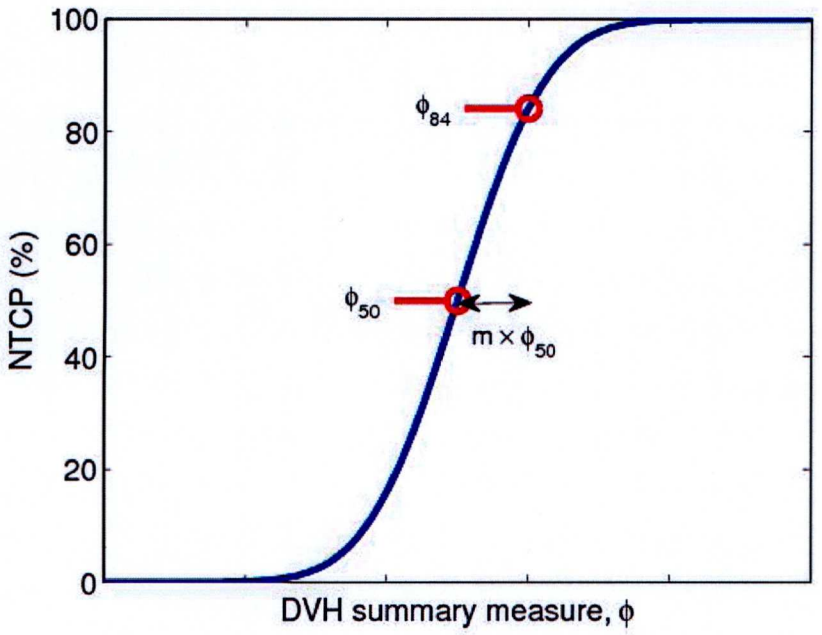


Figure 2.1: Illustration of the s-shaped curve given by the probit function with mean ϕ_{50} and relative standard deviation m.

In the presentations which follow, it is assumed that all doses have been normalised with respect to fraction size and expressed as EQD2, as described in section 1.3.

2.1.1 The LKB model

The DVH is most often reduced to the dose which would cause the same NTCP if given uniformly to the whole organ, the equivalent uniform dose (EUD). Different expressions for the EUD have been suggested, based on different functions of local dose effect. As shown by Niemierko [54] and Seppenwoolde *et al.* [43] EUD can be calculated by

$$\sum_i \mathrm{E}(D_i)\frac{V_i}{V_{tot}} = \mathrm{E}(\mathrm{EUD}) \quad \Leftrightarrow \quad \mathrm{EUD} = \mathrm{E}^{-1}\left(\sum_i \mathrm{E}(D_i)\frac{V_i}{V_{tot}}\right) \tag{2.3}$$

where $\mathrm{E}(D_i)$ is the local dose effect function. Consistent with Lyman's model [51], Mohan *et al.* [55] chose a power-law local effect function when deriving an 'effective dose'. Later Niemierko [56] used an equivalent expression for his generalised EUD[1]:

$$\mathrm{EUD} = \left(\sum_i D_i^{\frac{1}{n}}\frac{V_i}{V_{tot}}\right)^n \tag{2.4}$$

[1]Niemierko, however, used the parameter a instead of n, where $a = \frac{1}{n}$.

Other expressions have been suggested for the equivalent uniform dose (or similar concepts), such as one based on the LQ model [54] where equal cell kill is assumed to cause equal effect, and thus is most suitable for evaluating the dose distribution in the tumour, and one based on the relative seriality model (section 2.1.3) [57]. However, equation 2.4 is the one most commonly used, and hereafter EUD will always refer to Niemierko's generalised EUD.

The probit function combined with the EUD DVH reduction scheme has three parameters. Here ϕ = EUD and ϕ_{50} is generally called TD_{50} (c.f. figure 2.1). This model is equivalent to the LKB model which instead of EUD calculates an effective volume:

$$V_{eff} = \sum_i \left(\frac{D_i}{D_R}\right)^{1/n} \frac{V_i}{V_{tot}} \tag{2.5}$$

Here D_R is a reference dose, normally the maximum or prescription dose. V_{eff} is actually not a volume since it is dimensionless, but rather a volume-weighted dose-effect normalised to D_R. In this model ϕ_{50} appears to be represented by the 'partial volume' TD_{50}, normalised to D_R,

$$TD_{50PR} = TD_{50} \cdot V_{eff}^{-n}, \tag{2.6}$$

and ϕ by D_R, which implies that ϕ_{50} has a different value for different DVHs. However, if the expression for t (equation 2.2) is manipulated it can be shown that ϕ equals EUD and $\phi_{50} = TD_{50}$, and thus the LKB model is identical to the EUD model:

$$t = \frac{D_R - TD_{50PR}}{m \cdot TD_{50PR}} = \frac{D_R - TD_{50}\left(\sum_i \left(\frac{D_i}{D_R}\right)^{1/n} \frac{V_i}{V_{tot}}\right)^{-n}}{m \cdot TD_{50}\left(\sum_i \left(\frac{D_i}{D_R}\right)^{1/n} \frac{V_i}{V_{tot}}\right)^{-n}} =$$

$$= \frac{D_R\left(\sum_i \left(\frac{D_i}{D_R}\right)^{1/n} \frac{V_i}{V_{tot}}\right)^{n} - TD_{50}}{m \cdot TD_{50}} = \frac{\left(\sum_i D_i^{1/n} \frac{V_i}{V_{tot}}\right)^{n} - TD_{50}}{m \cdot TD_{50}} = \frac{\text{EUD} - TD_{50}}{m \cdot TD_{50}}$$

Thus, these models are the same, and when used it will be called the LKB model, because of its historical importance and current popularity, but generally the EUD formalism will be used, due to its greater transparency.

A two-parameter version of the LKB model is the mean dose model. Here $n = 1$ and EUD is equal to the mean dose:

$$\text{EUD} = \sum_i D_i \frac{V_i}{V_{tot}} \tag{2.7}$$

The difference in the local dose effect assumed by the LKB model and the mean dose model is illustrated in figure 2.2. For the mean dose model this is a straight line. The power-law local dose effect function used by the LKB model is an approximation of an s-shaped curve for low effects [58]. For some organs a high level of local tissue damage is acceptable if limited to a small volume, and in this case the linear or power-law approximations of the LKB DVH reduction schemes might not be the best choice. The versions of the critical volume model below offer alternative shapes of the local dose effect function.

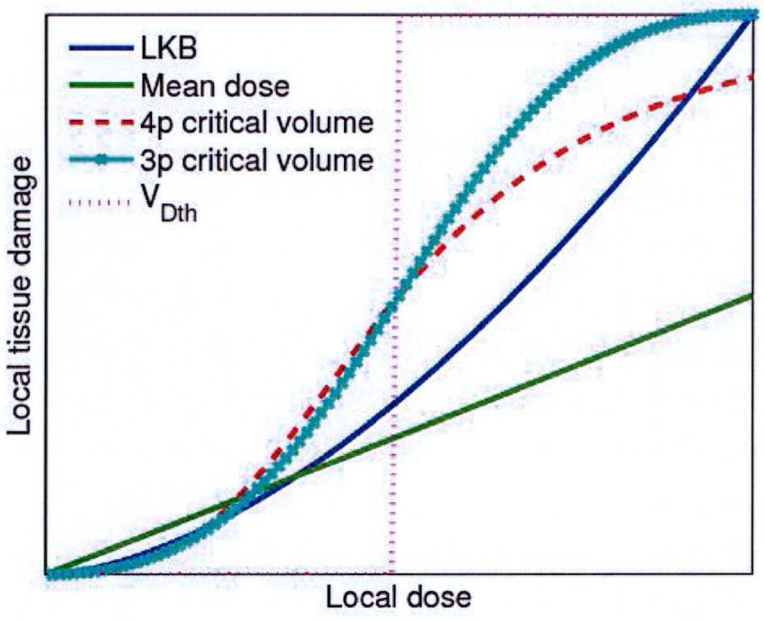


Figure 2.2: The shape of local dose effect functions assumed by the different DVH reduction schemes used by the LKB model, the mean dose model, the four-parameter critical volume model, the three-parameter critical volume model and the V_{Dth} model respectively.

2.1.2 The critical volume model

The critical volume model was developed as a biological model based on the LQ model and Poisson statistics (see section 2.2.1). This section will focus on a purely empirical

model used by the Dutch NKI group [59, 43], which resulted from a simplification of the original critical volume model. The local dose effect function (see figure 2.2) is

$$\mathrm{E(D_i)} = \frac{1}{1 + (D_{50}/D_i)^k}, \tag{2.8}$$

and the summary measure which can be used with the probit function is the relative damaged volume:

$$\phi = \sum_i \mathrm{E(D_i)} \frac{V_i}{V_{tot}}. \tag{2.9}$$

The model is named after the critical damaged volume parameter, ϕ_{50}. The other three parameters of the critical volume model are m, D_{50} and k.

When this model is fitted to clinical or experimental data it is often difficult to find a unique best estimate for both local effect parameters, especially the parameter representing the slope of the local dose-effect function, k [60, 43, 44]. Therefore, it is more efficient to assume that the slope of the curve is known, which reduces the model to a three-parameter model. Jin et al. [61] have proposed this one-parameter s-shaped function for the local dose effect:

$$\mathrm{E(D_i)} = \begin{cases} \frac{D_i/D_{50}-1}{1+(D_i/D_{50}-1)^2} + 1/2 & \text{when } D_i < 2D_{50} \\ 1 & \text{when } D_i \geq 2D_{50} \end{cases}, \tag{2.10}$$

which assumes a moderate slope and is symmetric around D_{50} (see figure 2.2). This function replaces equation 2.8 in equation 2.9.

Where a steep dose-effect function is more appropriate the V_{Dth} model is to be preferred. This model assumes that volumes receiving doses above D_{50} are completely damaged, whilst volumes with lower doses are spared. The local dose effect function (see figure 2.2) is:

$$\mathrm{E(D_i)} = \begin{cases} 0 & D_i < D_{50} \\ 1 & D_i \geq D_{50} \end{cases} \tag{2.11}$$

In this section three alternative local dose effect functions have been presented (equations 2.8, 2.10 and 2.11). All three version of the critical volume model can be fitted to a dataset and the one that performs the best be chosen (see section 3.3).

2.1.3 The relative seriality model

Unlike the models described above the relative seriality model [47] is not based on the probit function. The DVH is expressed as a probability-volume histogram (also called a risk histogram [62]):

$$P(D_i) = 2^{-\exp\left(e\gamma\left(1-\frac{D_i}{D_{50}}\right)\right)}, \tag{2.12}$$

where D_{50} is the dose causing 50% NTCP if the organ is uniformly irradiated, and γ is the slope at the steepest point of the curve. The NTCP is given by

$$NTCP = \left(1 - \prod_i (1 - P(D_i)^s)^{\frac{V_i}{V_{tot}}}\right)^{1/s}, \tag{2.13}$$

where s is the volume-effect parameter. For high values of s (~ 1) this function assumes that the FSUs are organised predominantly in a serial manner, whilst for low values of s a large number of FSUs are assumed to function independently of each other.

It will be shown in chapter 7 that the performance of the relative seriality model and the LKB model often are remarkably similar when fitted to the same dataset. This might be due to the features they have in common: a dose effect calculation based on knowledge about whole-organ uniform irradiation, and a power-law histogram reduction (probability-volume histogram and DVH, respectively).

2.2 Mechanistic and non-DVH based NTCP models

The models described in the previous section are general enough to be suitable for most clinical datasets, and though they can be interpreted to reflect certain assumptions about the underlying biology they are essentially empirical, and NTCP estimates can fill the same purpose as empirically derived dose-volume constraints [42]. A number of different models can be fitted to a clinical dataset and the one which performs the best (see section 3.3) can be selected for future treatment plan evaluation for example.

Since the DVH reduction schemes are empirical there is an uncertainty in the actual risk connected with a certain value of ϕ, a population standard deviation due to the different dose distributions giving similar ϕ but potentially associated with different NTCP. In some situations it is of interest to explore the potential of more specialised models, which more explicitly reflect the pathogenesis of a certain

complication. Typically, this is the case where non-local effects of irradiation cannot adequately be captured by an empirical DVH reduction scheme, such as non-local repair effects for very small fields of very high doses achieved by heavy ion irradiation [42], or the so far unexplained additive damage by volumes of very low doses to that caused by volumes receiving significant doses [63]. Here, the DVH reduction schemes fall short since the damage in organ subvolumes contribute independently to the total damage.

Another reason to explore mechanistic models is the potential to use them for untried treatment techiques. A model which adequately reflected the pathogenesis of a complication would be more reliable than an empirical model when extrapolating to dose distributions very different from the current clinical experience.

If a tissue consists of FSUs which are damaged independently by radiation and if organ function depends on the integrity of some or all of these FSUs, binomial statistics can be used to calculate NTCP. This approach is altogether different from the empirical use of the models described so far, where the main concern is to detect statistical patterns in a dataset. A model developed around the FSU concept aims to estimate the NTCP based on the knowledge of radiobiological mechanisms, independently of previously collected clinical data. Model parameter values can be chosen based on experimental measurements or reasoning about the organ architecture.

2.2.1 The mechanistic critical volume model

A model developed on these principles is the mechanistic critical volume model [62, 34, 64]. A functional reserve parameter decides how many FSUs can be damaged before a complication occurs; the probability that this threshold is exceeded is given by the cumulative binomial distribution (or the normal distribution if the number of FSUs is large); and the probability of FSU inactivation is given by the LQ model. The model was applied to the kidney and values for the number of stem cells per FSU, the stem cell radiosensitivity, the total number of FSUs and the functional reserve were suggested. With such a model also the level of intra- and inter-individual variations in the parameter values need to be taken into account, which leads to a multitude of model paramteters. Therefore, simplified versions of the model were suggested.

The implementation of the critical volume model in section 2.1.2 relates to the individual (rather than population) critical volume model with the approximation of

the binomial distribution with the normal distribution [34], but it is purely empirical since the parameter values when fitted to a population do not represent the biological parameters built into the mechanistic model [64].

2.2.2 Modelling spinal cord injury

A special case of the critical volume model is the critical element model, which includes no functional reserve. Here each FSU is crucial for continued organ function. The binomial expression for NTCP is simpler in this case, and also in the case applicable to the tumour where the response depends on the probability of killing every stem cell (in this case each stem cell can be regarded as an FSU) [62, 34, 64]. The expression for the critical element model is

$$NTCP = 1 - (1 - P_{FSU})^{N^{FSU}}, \tag{2.14}$$

where P_{FSU} is the probability of inactivating an FSU and N^{FSU} is the number of FSUs in the organ. If equation 2.12 is used for calculating P_{FSU}, this model is identical to the relative seriality model with $s = 1$. Alternatively equation 4.1 can be used.

The critical element model has mostly been applied to the spinal cord which is typically regarded as a serial organ. However, some experimental data suggest that this approach is not valid when small or non-contiguous volumes of spinal cord are irradiated, and several attempts have been made to explain these results with alternative models. Stavreva *et al.* [65] adapted the critical volume model to account for the contiguity of the FSUs, and van Luijk *et al.* [66] developed a model which includes the effect of tissue repair through cell migration.

2.2.3 Cluster models

The requirement that damaged FSUs be contiguous was also adopted by Thames *et al.* [67] in their cluster model. In this model a complication is assumed to occur depending on the size of the largest cluster of contiguously damaged FSUs. The model was applied to paralysis caused by spinal cord injury, and later to rectal bleeding [68]. The spinal cord was considered a one-dimensional array of FSUs and the rectal wall a two-dimensional array. The critical size of the cluster for rectal bleeding was explored by

fitting the model to clinical data with pairs of patients with similar DVHs but different outcome, in order to demonstrate an eventual spatial effect which would be lost in the DVHs.

Although Fenwick's [69] parallel model for rectal bleeding is a critical volume type model, it has many similarities to the cluster models. An s-shaped function is used to model the risk of local tissue damage (telangiectasia), and the NTCP is given by another s-shaped function based on the number of radiation-induced telangiectasia. However, this model does not take the spatial distribution of local damage into account.

2.2.4 The variable critical volume model

The critical volume tolerance method [70] and the variable critical volume model [71] also focus on the damaged rectal wall area, but without making assumptions about the actual tissue damage mechanism. The variable critical volume model identifies the volume with the greatest damage for different sizes of the subvolume, and the damage threshold for triggering a complication is higher the smaller the volume is. This is in fact an empirical model since it is fitted to datasets to determine parameter values.

2.3 Discussion

2.3.1 The role of NTCP modelling

As described in section 2.2, there are two different classes of NTCP models, which are useful for different purposes. When the purpose is treatment plan evaluation it is important that the NTCP prediction is reliable, and efficiently incorporates the clinical experience. Conformal treatment plans often generate highly inhomogeneous dose distributions in normal tissues and such data are too complex to evaluate manually [42]. NTCP models fitted to clinical data have been 'trained' to identify characteristics of DVHs which are associated with a higher risk of toxicity, and if used together with confidence intervals the reliability of the NTCP estimate is apparent (see section 3.2). An ideal empirical NTCP model should classify high- and low-risk treatment plans with high accuracy (see section 7.4.1). This will have an impact on the optimal number of parameters, which if too high will prevent the limited size of clinical datasets from determining the parameter values with adequate precision, and if too low will fail to

describe the data. Therefore empirical models have been developed to be fast and easy to use [66], and have a high degree of transparency [42].

As suggested by Withers and Taylor [72] a biologically based mechanistic model, on the other hand, is dependent on an extensive quantitative knowledge base of radiobiology. These models highlight what areas need further research before a full understanding of radiation-induced pathogenesis can lead to predictive modelling of NTCP, without the need of statistical fitting of empirical parameter values. They also demonstrate what can be achieved when such knowledge is available. Thus, mechanistic NTCP modelling is dependent on research in experimental radiobiology, but can also support and inspire such research by generating hypotheses and suggesting explanations. However, the applicability of these models to treatment planning is still limited, due to the many unresolved issues of normal-tissue response mechanisms [73]. Also, biologically based models are necessarily simplifications of reality, and it can be argued whether the strict assumptions involved lead them too far away from reality, and whether it is even possible to develop a comprehensive biologically based NTCP model [46].

2.3.2 The LQ model

When NTCP is calculated for inhomogeneous dose distributions the DVHs should always be normalised, e.g. to 2-Gy equivalent doses according to equation 1.3 [74]. Otherwise, if 'physical' total doses are used the contribution from damage to low dose volumes is overestimated, due to the non-linear nature of the relationship between dose and cell kill. Unlike the mechanistic NTCP models described in section 2.2, the LQ model used to normalise DVHs has been validated extensively on experimental data [16], which justifies considering the normalised doses as given, rather than part of the fitted NTCP model. Some investigators, though, have fitted the parameter α/β to their dataset together with the NTCP model parameters, when used with an empirical model. In this case the LQ model is also considered empirical. When used in a mechanistic sense, on the other hand, the value of α/β is chosen based on experimental results. This is not to say that α/β is generally very precisely known, but the choice of α/β seems to have weak influence on the fitted NTCP parameter values [75, 76]. To be on the safe side the calculations can be repeated for a few different values around the standard $\alpha/\beta = 3$ Gy for late-responding normal tissues and $\alpha/\beta = 10$ Gy for early-responding normal tissues and tumours [15].

However, care must be taken to use a version of the LQ model appropriate to the treatment technique. In its simplest form (equation 1.2), complete repair of sublethal damage between fractions and absence of proliferation are assumed. Thus, when applied to hyperfractionated treatments (if the interfraction interval is less than about six hours) and/or to early responding normal tissues, time effects must be included [77, 78]. Also, when a treatment takes longer than $15 - 30$ minutes, intra-fraction repair decreases the effect of the fraction dose and must be taken into account [77]. This is important when normalising DVHs for brachytherapy as well as time consuming conformal treatments like stereotactic body radiotherapy (SBRT), intensity modulated radiotherapy (IMRT) and tomotherapy. However, as pointed out by Stewart and Dörr [19], the influence of treatment time is still not well understood and it is likely that cell migration as well as repair of sublethal cell injury influence the tissue response over time [14]. Therefore, calculations depending on time effects should be treated with caution.

2.3.3 The volume effect

The rationale for efforts to conformalise radiotherapy is the hypothesis that reduced volumes of normal tissues irradiated to high doses are associated with lower complication probability. When the number of beams used to deliver the tumour dose is increased the normal-tissue volume receiving high doses can be decreased, but the larger number of beams leads to larger normal-tissue volumes being irradiated, though this dose is relatively low. The integral dose to the normal tissues around the target is approximately constant for different photon beam arrangements [79, 80], but the effect of increasingly inhomogeneous dose distributions is difficult to predict. When a traditional endpoint is considered, small volumes of organs with a large functional reserve might tolerate extremely high doses, but then a new type of toxicity might be encountered which has a different dose-volume dependence. Also, it might be considered safe to irradiate large volumes to low doses, but in addition to smaller volumes of high dose the damage in these volumes might not be negligible [79, 63]. The risk associated with different dose distribution characteristics remains an important field of research [81].

One way of addressing this issue is to ask whether a lot (of dose) to a little or a little to a lot is to be preferred. A study on lung treatment data concluded that dispersing the energy over large volumes results in a lower NTCP [82], whilst another analysis

showed that organs with a large functional reserve are likely to have a higher NTCP if treated with low doses to large volumes [80]. Serial organs, on the other hand, benefit from suppressing the highest doses rather than the volumes receiving low doses. In conclusion the question of a lot to a little or a little to a lot is over simplified. The answer depends both on the organ architecture and on how high the dose in the 'low' dose volume is [79, 82]. The response of tissues with radiosensitive FSUs would be expected to depend on lower doses, and for conformal treatments to be effective the 'low' dose given to large volumes would have to be low enough not to cause significant FSU damage [34].

A related issue is the existence of a volume threshold such that there is no limit to the dose which can be delivered to an organ fraction below this threshold. This is mostly a philosophical question introduced by differing extrapolations from different models [35], since there are practical limits to the dose that can be delivered to a subvolume without causing excessive damage elsewhere. However, with SBRT exceptionally high doses can be delivered with very low incidence of complications, which seems to support this hypothesis. Schultheiss [33] suggests that all organs have (endpoint specific) volume thresholds, but that in some cases these are too small to be clinically relevant.

Chapter 3

Statistical Methods for Model Fitting and Evaluation

3.1 The likelihood function

This section will focus on the methodology for fitting empirical NTCP models to datasets, and methods for evaluating the model fit are described in section 3.3. It is assumed that the data set consists of DVHs for the individual patients and binary type clinical outcome data, e.g. yes/no for grade 2 rectal bleeding.

3.1.1 Parameter estimation

When modelling binary data the maximum likelihood method is the parameter estimation method of choice [45, ch. 3]. The best estimates of the parameter values are calculated by maximising the likelihood of observing the actual outcome data for the corresponding dose distributions D_i [83, ch. 8]. Let $Q_i = 1$ for patients who suffer a complication, $Q_i = 0$ for patients without the complication, and $NTCP_i(D_i, \boldsymbol{a})$ be each patient's NTCP estimated for a given dose distribution and parameter set $\boldsymbol{a}$. Then the likelihood P of observing $\mathbf{Q}$ for $\mathbf{NTCP}$ is given by

$$P = \prod_i NTCP_i^{Q_i} \left(1 - NTCP_i\right)^{1-Q_i}, \tag{3.1}$$

where $i = (1, 2, ..., N)$ are the patients in the dataset. The task of finding the optimal parameters then consists of maximising equation 3.1 in terms of the parameter set $\boldsymbol{a}$,

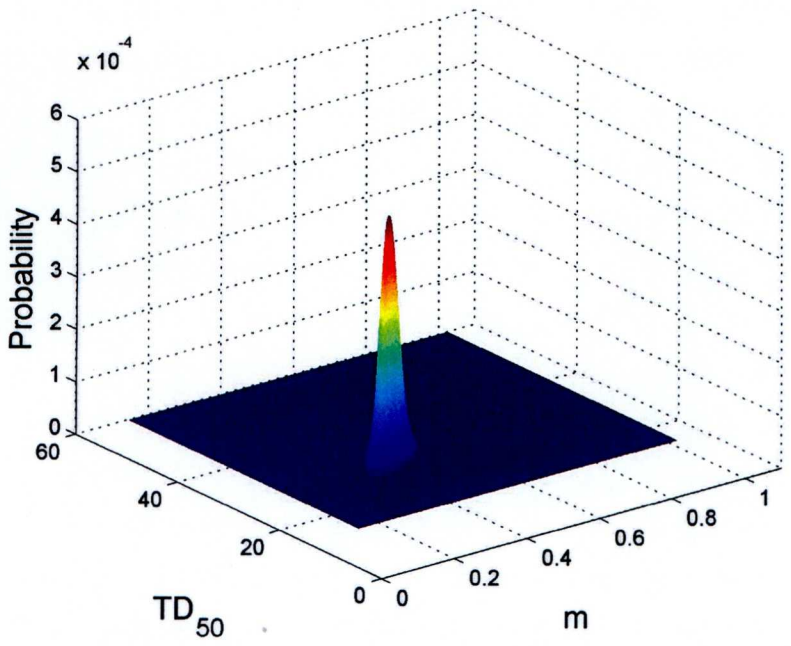


Figure 3.1: An example of a probability distribution for the mean dose model.

or equivalently the log of equation 3.1:

$$L = \sum_i Q_i \ln\left(NTCP_i\right) + (1 - Q_i)\ln\left(1 - NTCP_i\right) \tag{3.2}$$

The maximum likelihood parameter estimate is a point in the full probability distribution. Figure 3.1 shows this probability distribution for the two-parameter mean dose model (section 2.1.1). This probability surface is created by the likelihood function 3.1 for all the combinations of parameter values in the chosen parameter space, normalised by dividing the distribution by its integral [84, 85]. The maximum likelihood estimate of the parameter values is at the peak.

An alternative choice of parameter estimate is the central interval estimate which is calculated by projecting the probability distribution onto each parameter axis in turn (figure 3.2), and selecting the central parameter value[1]. Since the probability function generally is not symmetric the maximum likelihood and the central interval estimates are different. A central interval estimate of the confidence interval (CI) of the parameter estimate can also be derived from the projection of the probability distribution; see figure 3.2. The 68% CI is given by the parameter values where the integral of the function below and above the values respectively is 16% [85].

[1]The central value is the parameter value for which the area under the curve is equal on either side.

If the maximum likelihood method is used without calculating the full probability distribution it is important to use an optimisation algorithm which converges to the global minimum, e.g. simulated annealing, since many datasets produce probability distributions with more than one local minimum.

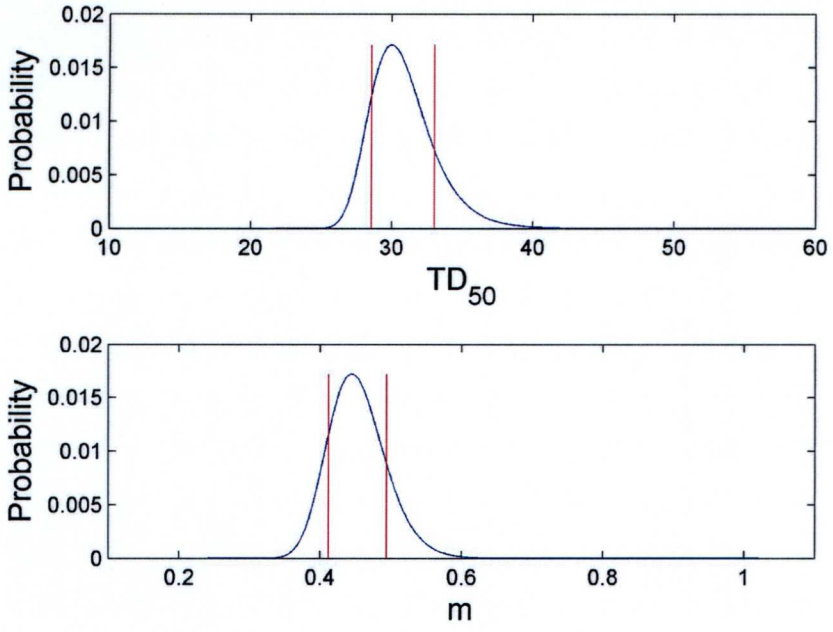


Figure 3.2: Projection of the probability distribution in figure 3.1 onto each of the parameter axes, from which the 68% confidence intervals can be calculated.

3.1.2 NTCP estimate confidence intervals

The probability distribution in figure 3.1 can also be used when estimating the NTCP with CIs for a given DVH. The NTCP for this DVH is then calculated for the same parameter combinations as for the probability distribution; see example in figure 3.3. Further, a histogram is calculated from the NTCP distribution, weighted by the corresponding likelihood values in the probability distribution. The resulting NTCP histogram in figure 3.4 gives the NTCP estimate with CIs for the DVH which figure 3.3 corresponds to.

Thus, the probability distribution is created based on an appropriate dataset with clinical and dosimetric data, and when used to estimate the NTCP for a given DVH a unique NTCP distribution is constructed, and the probability- and NTCP-distributions together yield a DVH-specific NTCP histogram.

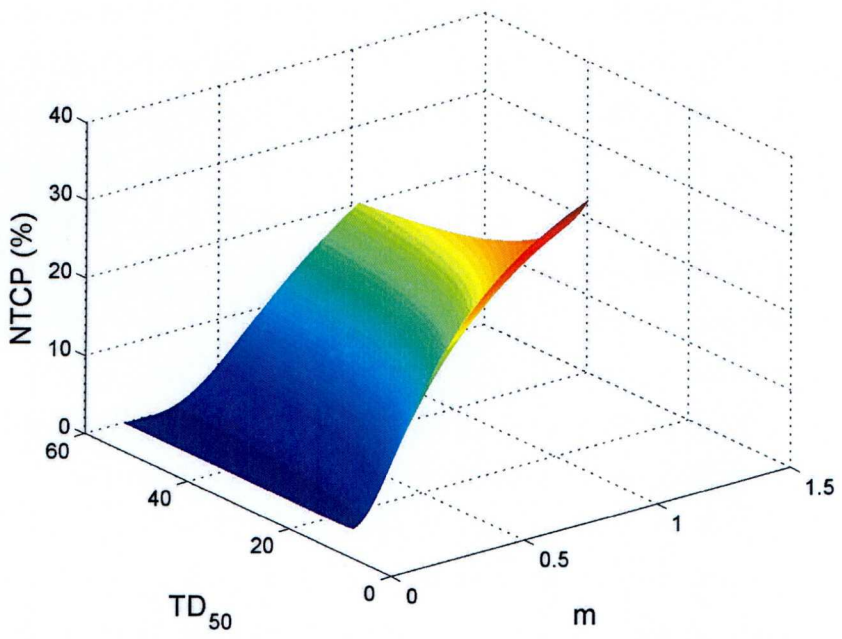


Figure 3.3: NTCP distribution for a DVH, for many alternative combinations of model parameter values.

The NTCP estimate in figure 3.4 is based on the full probability distribution for the model parameter values, not only the maximum likelihood estimates. The confidence interval expresses how certain this estimate is, based on the clinical experience collected in the dataset used to build the probability distribution. This means that if, hypothetically, many patients received dose distributions to the organ at risk giving exactly the same DVHs, and divided into cohorts, the complication rate in 19 out of 20 of these cohorts would fall within the 95% CI.

The width of the CI is influenced by lack of information on several levels. Firstly, if the model is inherently unsuitable for summarising the collected dataset (e.g. too few parameters, unrealistic local dose effect function) the parameter estimates will have large uncertainties, which translate into the NTCP estimate CIs. Secondly, since the model is fitted only to dose-volume data, confounding factors, such as inter-patient variation in radiosensitivity and health status, always decrease the power of a model to describe the observed clinical outcome (assuming no patient-specific information except the DVH is included in the model), resulting in larger CIs. Thirdly, also the loss of information in the reduction of the dose distribution to a DVH is a potential confounding factor. Fourthly, the parameter CIs are large if there are limitations in the clinical experience due to the dataset being too small or containing very similar DVHs.

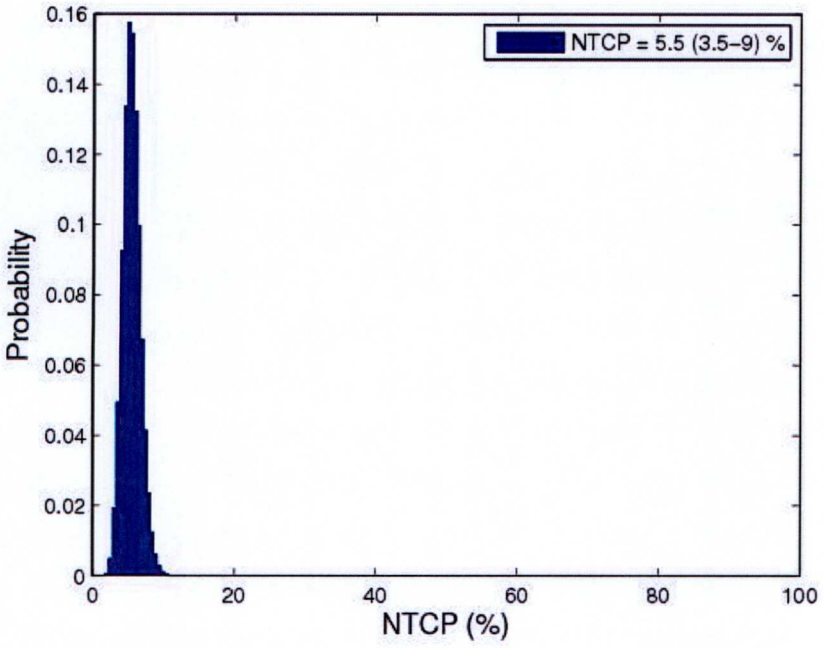


Figure 3.4: NTCP histogram giving an NTCP estimate with confidence intervals for a given dose distribution, based on previously collected clinical data.

If the model has an appropriate summary measure it should yield large CIs for a DVH which is very different from the DVHs in the dataset used to create the probability distribution [85, 43]. For example, if the LKB model is fitted to a dataset with close to uniform irradiation of the organ at risk, there will be little experience of the volume effect of the organ in this dataset, and the volume parameter n will have a large CI. This will result in larger CIs for a new patient with an inhomogeneous dose distribution than a patient with a uniform one. In this way new treatment techniques could be explored by simply making sure the upper limit of the NTCP CI for patients doesn't exceed a clinical tolerance level. However, since it is unlikely that the DVH-based models extract all the important dose-volume-response information from a dataset, such faith in the CIs is not justified. This is confirmed by a comparison of NTCP estimates with CIs for different models, based on the same dataset and applied to the same DVH but where the CIs do not overlap, as shown in section 7.4.1. Therefore it is important that users understand the mechanisms of the NTCP model and what features of the dose distributions cannot be captured by the model [42].

3.2 Application: individualised dose prescription

The following analysis [86] highlights the importance of the use of NTCP confidence intervals.

The clinical use of complication probabilities derived from NTCP models is limited by the uncertainties in the model parameters, which are derived from responses observed in a population of patients. Associated CIs would give the clinician valuable information about the uncertainty in the NTCP prediction, for a given dose plan and prescription dose. The aim of the work described in this section is to show how such confidence intervals can be derived from the literature or from local clinical data, and used together with estimates of TCP to facilitate individualised dose-prescription.

DVHs from 31 patients receiving lung tumour radiotherapy at Clatterbridge Centre for Oncology (CCO) were used to illustrate the potential of including NTCP CIs in individualised treatment planning. All DVHs were normalised to 2-Gy fractions according to equation 1.3, using a value of $\alpha/\beta = 3$ Gy. The data used to build NTCP parameter probability distributions were mean lung dose (MLD) and radiation pneumonitis data from the literature; see below. Since the dose-volume data available was MLD rather than DVHs, two-parameter versions of the LKB (mean dose model; see section 2.1.1) and relative seriality models were used. When $D_i = \text{MLD}$ for all i the relative seriality NTCP is given by equation 2.12 and is independent of s. Similarly, when applying the fitted models to the CCO DVHs for lung minus gross tumour volume (GTV), the DVHs were reduced to the MLD and the two-parameter models were used to calculate NTCP.

TCP was calculated for the planning target volume (PTV) DVHs with the Marsden model [87]:

$$TCP_{pop} = \frac{1}{\sigma_\alpha \sqrt{2\pi}} \int_0^{\inf} TCP(\alpha) \mathrm{e}^{-\frac{\alpha-\overline{\alpha}}{2\sigma_\alpha^2}} \, \mathrm{d}\alpha, \tag{3.3}$$

where, for a given α

$$TCP(\alpha) = \prod_i \mathrm{e}^{-\rho v_i \mathrm{e}^{\alpha n d \left(1+\frac{d_i}{\alpha/\beta}\right) + \ln 2 \frac{t_{treat}-t_{lag}}{T_{eff}}}}. \tag{3.4}$$

Equation 3.3 gives the average TCP for a population with different values of α. The following parameter values were used: mean radiosensitivity, $\overline{\alpha} = 0.31$ Gy^{-1} with

standard deviation $\sigma_\alpha = 0.062$ Gy^{-1}, $\alpha/\beta = 10$ Gy, clonogen density $\rho = 10^7$ cm^{-3}, clonogen doubling time $T_{eff} = 3$ days, and time to proliferation start $t_{lag} = 21$ days [88]. The total treatment time t_{treat} was assumed to be 28 days.

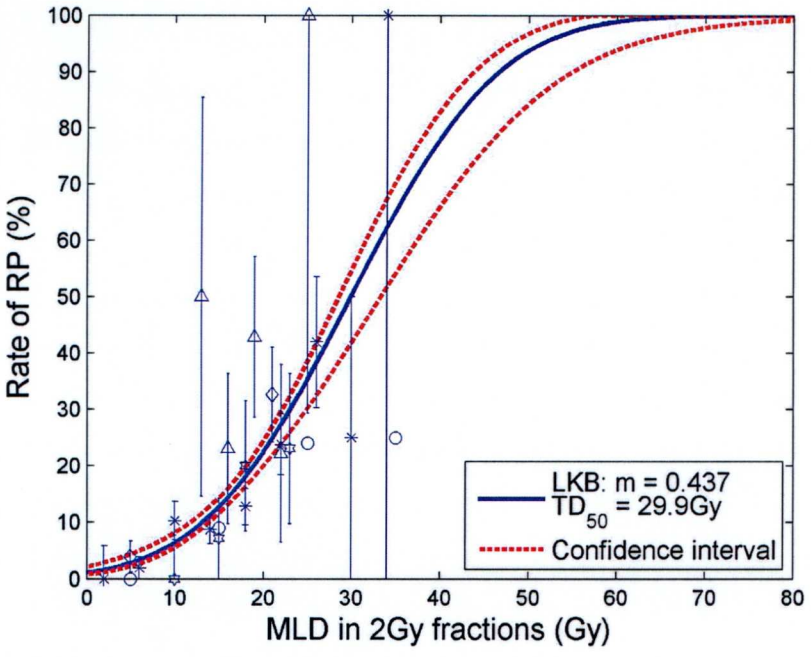


Figure 3.5: Grade ≥ 2 radiation pneumonitis rates for different MLD [89, 90, 91, 92, 82], with corresponding Poisson uncertainty estimates. Dose-response curve from maximum likelihood estimates of LKB parameters TD_{50} and m ($n = 1$ for MLD), and Monte Carlo 68% confidence intervals. The Monte Carlo confidence intervals are based on 5000 sets of parameters.

NTCP parameters were fitted to clinical data for grade-2 radiation pneumonitis (RP) versus MLD collected from the literature [89, 90, 91, 92, 82], first using the maximum likelihood method, as in figure 3.5 for the LKB model. In these publications patients had been binned according to MLD, so all patients in one bin were assumed to have the same MLD, and the number of responders in the bin was given by the corresponding complication rate. In figure 3.5 the 68% confidence interval for the fitted curve was derived using a Monte Carlo method, by randomly sampling 5000 artificial datasets from the real dataset, assuming a normal distribution of the uncertainty in the data points.

The probability distribution (PD) method (section 3.1) was used to calculate NTCP with CI for the individual CCO DVHs (for lung-GTV). Also TCP and complication-free control (CFC) were calculated (for the GTV DVHs), for a range of possible prescription doses; see the result for one patient in figure 3.6. CFC equals TCP(1−NTCP) and estimates the probability of a 'successful' treatment outcome.

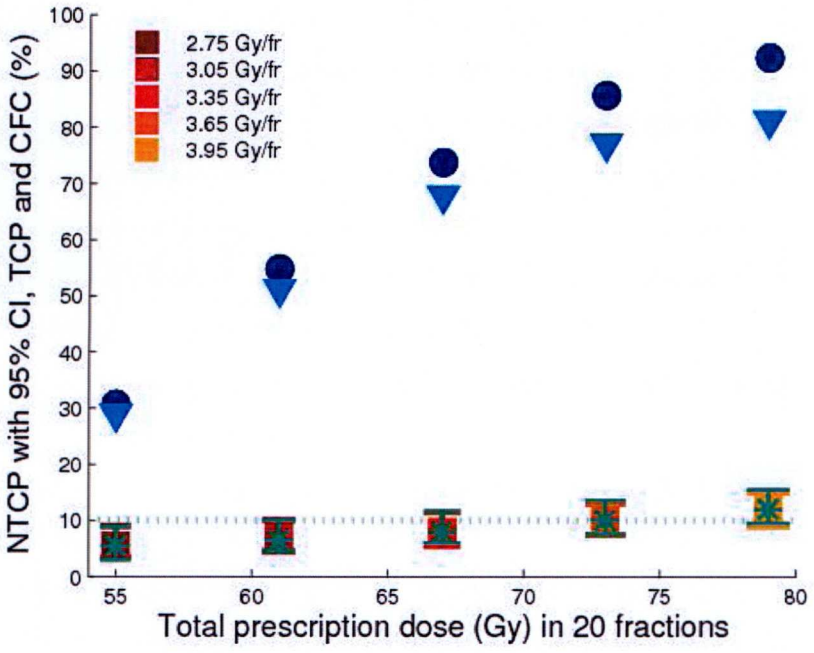


Figure 3.6: LKB (□) and relative seriality (∗) NTCP with 95% confidence intervals, compared to an accepted level of 10% grade ≥ 2 radiation pneumonitis, TCP (∘) and complication-free control (▽) for a typical patient, for a range of prescription doses.

The PD method gave the following LKB model parameters: $TD_{50} = 30.5$ Gy (68% CI: 28.6-33.1 Gy) and $m = 0.449$ (0.412-0.494), and for the relative seriality model $D_{50} = 32.7$ Gy (30.4-35.8 Gy) and $\gamma = 0.722$ (0.668-0.779). These estimates are similar to published parameter estimates from clinical studies, e.g. [93, 43]. This indicates that the PD method based on MLD data from the literature is consistent with other fitting methods.

Ideally, if enough local clinical data are available, the NTCP with CI can be calculated with a three-parameter PD, based on DVHs instead of MLD as input data. As more local outcome data is accumulated, the CIs on NTCP estimates will decrease and, in principle, be tailored to the treatment techniques of the centre concerned.

In a study about the importance of NTCP CIs for treatment planning Langer concludes that "the effect on tumour dose of relaxing a constraint within its range of uncertainty should be made available to physicians involved in the selection of treatment plans" [46]. Together with an estimate of TCP, the NTCP for a range of prescription doses can be a guide for the clinician in finding the optimal dose, as a trade-off between local control and complication probabilities. The CI of the NTCP estimates give an indication of the uncertainty in the (population averaged) predicted response of the patient, for a given dose plan and prescription dose.

Currently at CCO, for the standard 55 Gy to the tumour, the patient in figure 3.6 would achieve (on the average) NTCP = 5.5% and TCP = 31%. However, with individualised dose-prescription she/he might receive 73 Gy by accepting mean NTCP = 10%, resulting in a significantly increased TCP (86%). Alternatively, 61 Gy may be prescribed if the dose was conservatively based on the upper limit of the CI not exceeding NTCP = 10% (TCP = 55% in this case). The individualised prescribed dose may also be influenced by the health status of the individual patient, taking co-morbidities into account.

3.3 Model accuracy

After fitting a model to a dataset it is important to assess the accuracy, or goodness-of-fit (GOF), of the model. For a continuous response variable the model accuracy depends on the residual variance, but this approach cannot be applied to binary response variables since in this case the aim of the model fitting is not to draw a curve as close as possible to data points.

Depending on the purpose of the NTCP model different methods for assessing the accuracy of the model fit have been used. Here only the accuracy of models with predictive purposes will be discussed. In this case the purpose of the assessment is to estimate the strength of correlation between the predicted NTCP and outcome, or the power of the model to correctly separate responders from non-responders.

If the purpose of the assessment of accuracy is mainly to compare the performance of different models, only a relative GOF is necessary. However, if the models have different numbers of parameters it is important to consider the principle of parsimony; models with many parameters are generally able to fit datasets better, but this is of no use if the model is overfitted. In this case the model interprets spurious effects as universal and when applied to a different dataset it will perform poorly.

3.3.1 ROC analysis

A method which has been shown to be useful for assessing the predictive ability of a dose-volume metric [94] is receiver operating characteristic (ROC) curve analysis [95, ch. 2]. This method has also been recommended for NTCP analyses by the QUANTEC study [96].

To create the ROC curve the true positive ratio (TPR) and the false positive ratio (FPR) are calculated for a series of NTCP cut-offs. The individuals in the dataset are labeled positives and negatives if their calculated NTCP is above or below the cut-off, respectively. The positives who actually suffered a complication are true positives and those who did not are false positives. Similarly, the negatives are either true or false negatives. The TPR, or sensitivity, is the number of true positives divided by the number of actual responders, and the FPR is the number of false positives divided by the number of actual non-responders. $1 - \text{FPR}$ is called the specificity.

The values of the TPR and FPR depend on the cut-off, and the ROC curve is the relationship between FPR and TPR as the cut-off is decreased in small steps from NTCP=100% to 0%; see figure 3.7. At a cut-off of 100% no individuals are labeled positives, so TPR=FPR=0. As the cut-off is decreased increasing numbers of individuals will be accurately labeled positives, resulting in a growing TPR. At some point the first false positive will be identified and then FPR will also start growing. The better the responders and non-responders are separated by the model, the closer the curve will tend to the top left corner. If the ROC curve follows the diagonal the model is no better than a random labeling of positives and negatives. It is therefore logical to use the area under the curve (AUC) as a measure of model accuracy.

The AUC is not only an intuitive measure of model accuracy but also, in fact, the probability that a randomly chosen responder has a higher NTCP than a randomly chosen non-responder [95, ch. 2]. Moreover, when the AUC is calculated with the trapezoidal rule it is also identical to the two-sample Mann-Whitney U-statistic [97]. As a general rule, a model with AUC>0.9 has a high accuracy, a model with an AUC of 0.7-0.9 has moderate accuracy and a model with an AUC of 0.5-0.7 has low accuracy [98].

Since the AUC is equal to the Mann-Whitney statistic it can be calculated as follows, rather than from a plotted curve [97, 95, ch. 4]:

$$\hat{A} = \frac{1}{n_1 n_0} \sum_{j=1}^{n_0} \sum_{i=1}^{n_1} \psi(X_i, Y_j) \tag{3.5}$$

Here $\hat{A}$ denotes the AUC for a given fitted model, n_1 and n_0 are the number of responders and non-responders, X_i the NTCP estimate for responder i, Y_j the NTCP estimate for

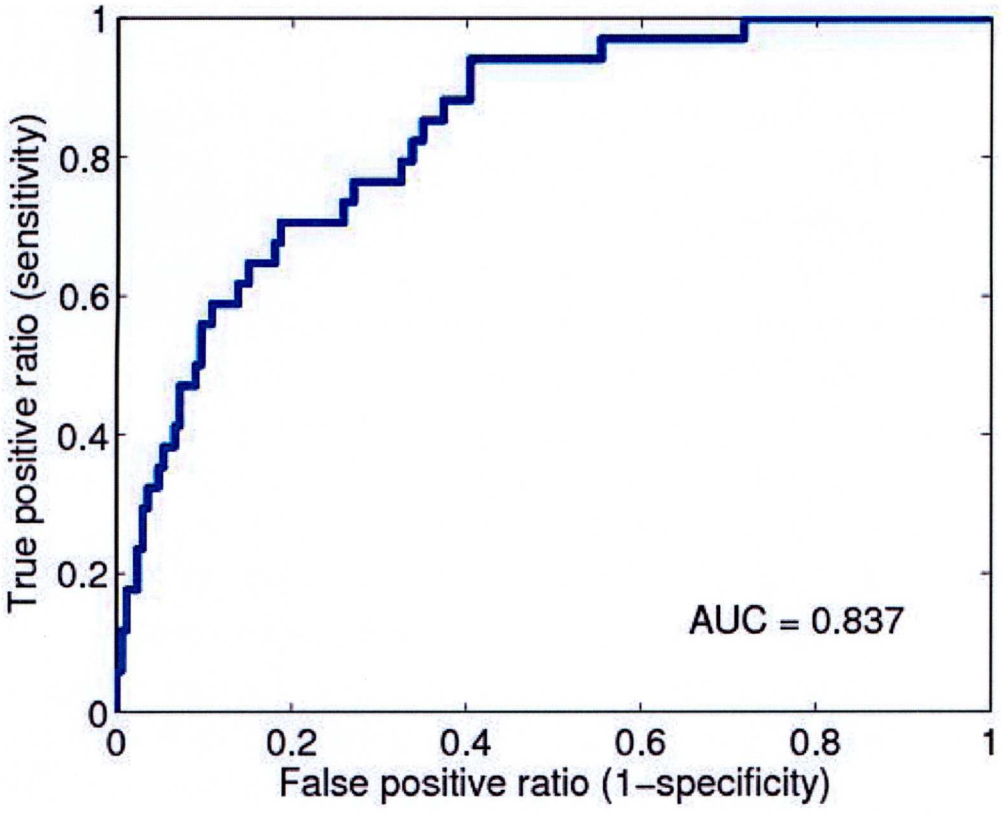


Figure 3.7: ROC curve for a dataset with 200 individuals. The area under the curve was calculated with the trapezoidal rule.

non-responder j, and

$$\psi(X,Y) = \begin{cases} 1 & Y < X \\ 0 & Y > X \end{cases} .$$

In order to calculate confidence intervals for the AUC and test for significance of a difference between the AUC for different models the covariance matrix $\mathbf{C}$ needs to be calculated. If k models are compared $\mathbf{C}$ is a $k \times k$ matrix. According to the method of DeLong *et al.* [97, 95, ch. 4-5] the following variance components are defined for model r:

$$V^r(X_i) = \frac{1}{n_0} \sum_{j=1}^{n_0} \psi(X_i^r, Y_j^r) \qquad (i = 1, 2, ..., n_1)$$

$$V^r(Y_j) = \frac{1}{n_1} \sum_{i=1}^{n_1} \psi(X_i^r, Y_j^r) \qquad (j = 1, 2, ..., n_0)$$

Also, the elements of $k \times k$ matrices $\mathbf{C}_X$ and $\mathbf{C}_Y$ are

$$c_X^{r,s} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (V^r(X_i) - \hat{A}^r)(V^s(X_i) - \hat{A}^s)$$

and

$$c_Y^{r,s} = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (V^r(Y_j) - \hat{A}^r)(V^s(Y_j) - \hat{A}^s).$$

Then the covariance matrix is given by

$$\mathbf{C} = \frac{1}{n_1}\mathbf{C}_X + \frac{1}{n_0}\mathbf{C}_Y. \tag{3.6}$$

The 68% confidence interval for A^r is calculated from the square root of the variance which is given by $\mathbf{C}_{r,r}$.

$\hat{A}$, given by equation 3.5, is an estimate of the 'true' AUC, A, given by an infinitely large dataset. To detect a statistically significant pairwise difference between the AUCs in $\boldsymbol{A}$ the following chi-square distributed test statistic is calculated:

$$Z = (\hat{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{L}'(\boldsymbol{L}\boldsymbol{C}\boldsymbol{L})^{-1}\boldsymbol{L}(\hat{\boldsymbol{A}} - \boldsymbol{A})' \tag{3.7}$$

The vector $\boldsymbol{L}$ is called the contrast and indicates which models are being compared. If e.g. $k = 4$ and model 1 and 3 are compared, $\boldsymbol{L} = (1 \quad 0 \quad -1 \quad 0)$. When two models are compared the test has one degree of freedom and a p-value can be calculated from the chi-square distribution for Z. Conveniently the unknown values in vector $\boldsymbol{A}$ always cancel out, since the sum of the contrast vector is 0.

The use of AUC as a GOF measure is convenient since it addresses the specific question of how well a model can be expected to predict the risk of toxicity, and furthermore the actual value has a relevant interpretation as the model's ability to correctly classify individuals as high- or low-risk cases. It should be noted however that the AUC is dataset dependent. Consider a set of DVHs which largely fall into two risk groups (e.g. if typically a very small or very large volume of normal tissue was irradiated). In this case several models might easily correctly classify responders and non-responders, whilst for a dataset containing many DVHs with NTCP $\sim 50\%$ all models have a greater challenge. This will lead to lower values of AUC. It is also important to note that the AUC is independent of the number of model parameters and does not safeguard against overfitting.

3.3.2 The Akaike information criterion

If the only concern is the relative performance of a set of NTCP models, a convenient GOF measure is the value of the maximum log-likelihood. The model which makes the observed outcome the most probable is the best model. However, like the AUC the maximum log-likelihood is vulnerable to overfitting. Instead it is better to use the Akaike information criterion (AIC) [99, ch. 2], which is based on the maximum log-likelihood (L) but gives advantage to models with a low number of parameters (N_{par}):

$$\mathrm{AIC} = -2L + 2N_{par} \tag{3.8}$$

Calculating AIC is asymptotically equivalent to cross-validating the model [99] on data subsets. However, the use of AIC is only valid for large datasets [99, ch. 2]. This method, and similar information criterion methods, have recently become popular for comparing the performance of NTCP models, e.g. [100, 101, 102].

3.3.3 The bootstrap method

An empirical bootstrap technique can be used to compare the GOF of different models [103]. With this technique many replicate datasets of the same size as the actual dataset are generated, by randomly sampling data points with replacement from the original data [100]. The model is fitted to each of the replicated datasets and confidence intervals for the original model parameters and the log-likelihood can be calculated. If this is done for two different models an empirical probability distribution of the difference between the log-likelihoods is easily derived. A significance test based on this probability distribution then shows if the difference between the log-likelihoods for the original dataset is statistically significant [75]. A disadvantage of this method is that it is computationally expensive.

3.3.4 Other goodness-of-fit methods

A GOF method used by Jackson *et al.* [104] has been very popular. They assumed normality of the log-likelihood function in order to estimate the expected maximum value of L and its standard deviation (see appendix in [104]). The probability of finding a lower value of L (and thus a worse fit), for an equivalent dataset, is given by comparing

the observed value of L to the normal distribution with the calculated expected value of L and its standard deviation as mean and standard deviation respectively. Thus, this method checks whether the non-random features influencing the outcome of the patients are adequately accounted for by the model prediction. However, the usefulness of this method is limited by its assumption that the value of L has a normal probability distribution.

The GOF method used by van Luijk *et al.* [60], the likelihood ratio test, is also based on the likelihood function but does not assume normality of the probability distribution. However, since this method operates on groups of identically treated individuals [45], it is more appropriate for experimental than clinical data. Similarly, Spearman's rank correlation, which has been advocated for NTCP modelling [105, 46], is not applicable to binary outcome data but can be used when the outcome is graded.

Two methodology and/or dataset dependent measures of GOF, which are suitable to use with binary data and do not make assumptions about the shape of the probability distribution, are the Hosmer-Lemeshow statistic and pseudo R^2 metrics [45]. The Hosmer-Lemeshow statistic is a chi-square statistic based on the observed and the model predicted complication rates in groups of individuals belonging to bins of predicted NTCP. Since the value of this statistic depends on how the data are binned the result is indicative rather than absolute.

A pseudo R^2 is a measure of the variation in the response explained by the model, and it ranges from 0 for a very poor model to 1 for an ideal model. The most popular pseudo R^2 was proposed by Nagelkerke [106]:

$$R^2 = \frac{1 - \left(\frac{M_0}{M(\boldsymbol{a})}\right)^{2/N}}{1 - M_0^{2/N}}, \tag{3.9}$$

where $M(\boldsymbol{a})$ is the value of the likelihood function for the fitted parameter values (the maximum likelihood), M_0 is the maximum likelihood for a model with a constant parameter only (called the intercept model for logistic regression), and N is the number of individuals in the dataset. In this way the model fit is compared to a poor model fit, and the larger the difference, the better the fitted model. The pseudo R^2 cannot be interpreted independently or compared for different datasets, but is useful for comparing the performance of different models for the same dataset. It was used by Rutkowska *et*

al. [107] for a logistic regression analysis in order to assess the strength of correlation between dose-volume parameters and outcome.

Chapter 4

A 3D Mechanistic Computer Model of Normal-Tissue Effects after Radiotherapy

The aim of this work was to develop a mechanistic model of normal tissue that can be used for simulating radiotherapy treatments. This chapter describes this model in detail. The 'patient', or organ, is represented by a 3D array of volume elements (voxels) containing FSUs, as well as doses representing the dose plan. Thus the dose in a voxel is assumed to be uniform, and the number of FSUs recorded in the voxel after 'treatment' depends on this dose and on the number of FSUs in the voxel before treatment. The model is based on two assumptions relating to overall organ injury and local tissue damage, respectively.

4.1 Organ injury

The first assumption concerns the organ response to local tissue damage after irradiation. It is assumed that every organ has a critical functioning fractional volume (CFV), which is the smallest fraction of the organ that, if its function is lost, causes a complication. This loss of function is assumed to occur when the number of surviving FSUs in the CFV is lower than a certain threshold value (FSU_{min}). In other words, in order for the organ to retain its function the number of surviving FSUs in any organ subvolume may not fall below FSU_{min}; see illustration in figure 4.1.

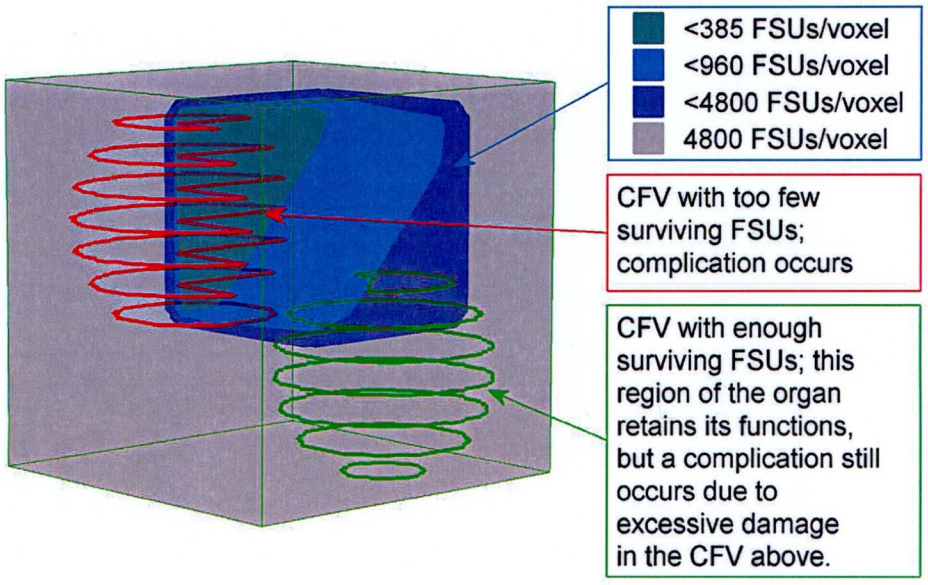


Figure 4.1: The 3D model represents an organ by an array of FSUs. Here the density of surviving FSUs per voxel is illustrated by the colourwash. The CFV (indicated by the wire-frame regions) here has a size corresponding to a semi-parallel organ; it can take any position in the normal tissue, but does not extend into the GTV. If a CFV in any position includes too few surviving FSUs (i.e. below FSU_{min}), a complication is triggered.

As a consequence, an organ with a small CFV (equivalent to *serial* tissue architecture) can suffer a complication even for a relatively small volume of severe tissue damage, whilst a large CFV (*parallel* architecture) allows organ function to be maintained despite damage to moderate volumes. The CFV can take any value above zero and less than or equal to 1, thus modelling the volume effect in a continuous fashion over a range of organ architectures; as the CFV goes from being very small for a serial organ to large for a parallel organ, it takes account of damage over an increasing number of contiguous voxels. Moreover, this volume effect representation also takes into account the spatial distribution of tissue damage when deciding whether a complication occurs. For example, in an organ with an intermediate-sized CFV, two small dose hotspots might cause a complication, depending on the distance between them. The model can also simulate the effect of an initially inhomogeneous density of FSUs, which could be caused by the natural tissue architecture, previous organ damage or the presence of a tumour, though this possibility is not explored in this thesis.

4.2 Local tissue damage

The second assumption defines the local effect of radiation on the tissue as a stochastic inactivation of FSUs. Here an FSU is defined as the largest tissue volume of damaged cells that can be regenerated by a single stem cell (see section 1.1.3). In some organs this might be the smallest unit performing a function in the tissue in question, e.g. if cells are prevented from migrating between such anatomical structures [29].

If complete repair of sub-lethal damage between fractions and absence of proliferation of normal tissues during the course of treatment are assumed, the stem cell surviving fraction is given by equation 1.2. As discussed in section 2.3, the normal-tissue endpoints considered in this case are late effects, which are caused by depletion of slowly proliferating stem cells, and the dose is assumed to be delivered at the rate of one fraction per day. For rapidly proliferating normal-tissue stem cells and/or hyperfractionated (i.e. twice a day) treatments, time effects can easily be incorporated into the model, however, by using the time-dependent version of the LQ model [78].

The probability of inactivating one FSU (P_{FSU}) is equal to the probability of lethally damaging every stem cell in the FSU, given by Poisson statistics if the number of stem cells per FSU, N_{sc}, is large (analogous to TCP modelling [108]):

$$P_{\mathrm{FSU}} = \mathrm{e}^{-N_{sc}\cdot \mathrm{SF}} \tag{4.1}$$

The number of surviving FSUs (N_S^{FSU}) in a voxel is stochastic, and therefore sampled from a binomial probability distribution, given P_{FSU} and the number of pre-treatment FSUs per voxel (N_0^{FSU}). However, for large N_0^{FSU} such that the binomial standard deviation $\leq 1\%$, the number of surviving FSUs is well approximated by the binomial mean and will be assumed to be deterministic:

$$N_S^{\mathrm{FSU}} = N_0^{\mathrm{FSU}}(1 - P_{\mathrm{FSU}}) \tag{4.2}$$

A similar approach has previously been taken by several other authors modelling NTCP mechanistically [62, 109, 34, 64].

Thus, equation 4.2 (or the randomly sampled N_S^{FSU}) applied to every voxel in the dose distribution gives a distribution of local tissue damage. Organ injury, however,

also depends on the volume effect which is modelled by the CFV. A complication is triggered if the number of surviving FSUs falls below FSU_{min} in any CFV.

4.3 Simulations

4.3.1 Model structure

A Matlab (www.mathworks.com) script was developed which simulates radiotherapy treatments as described above. Each step in the calculations is illustrated in the flowchart at the end of this chapter. First, values for input parameters are chosen; the organ specific parameters are the size of the CFV (critical functioning volume), FSU_{min} (threshold for the number of surviving FSUs in a CFV), the LQ model parameters α and α/β, N_{sc} (the number of stem cells per FSU) and number of FSUs/voxel. When data including inter-patient variation in radiosensitivity and health status are simulated, a mean and standard deviation for α and the number of FSUs/voxel are chosen and for each treatment these parameters are randomly sampled and recorded. The treatment specific parameters are dose per fraction and number of fractions.

Next the 3D array of a dose distribution is loaded. A set of such arrays was usually prepared in advance, as described in section 4.3.4, but could also imported from a treatment planning system. The doses are converted to EQD2 according to the selected treatment specific parameters. A PTV is defined based on the dose distribution as described in section 4.3.4, and a GTV is defined based on the PTV. A DVH and other dose-volume parameters for the normal-tissue dose distribution are calculated and recorded.

An array for the FSU distribution is created, which is of the same size as the dose distribution array. Since in the model local tissue damage is represented by FSU inactivation, this array initially represents the healthy normal tissue, and after irradiation the damaged normal tissue. In the FSU distribution, elements not belonging to the normal tissue of interest are set to zero (outside the organ and in the GTV), and within the organ it is uniform in its healthy state in the work presented in this thesis, although the model easily handles inhomogeneous FSU distributions. The distribution of surviving FSUs after applying the dose distribution is calculated on a voxel basis with the equations in section 4.2.

Next the number of FSU surviving in all possible CFVs throughout the organ should be calculated. When the model was developed, a few different approaches were tried and discarded before a satisfactory one was found. The main difficulty was to get the CFV to adapt its shape around the organ surfaces appropriately.

4.3.2 Modelling the CFV

In the first approach the CFV was given the shape of a cube, which was built by extending it equally in all three dimensions from one of its corners. Thus the number of CFVs fitting along each axis of the organ array was the organ array side length, minus the CFV side length, plus one. For any CFV including GTV voxels, layers were added to the CFV in each dimension such that the number of added voxels was roughly equal to the voxels initially shared by the CFV and GTV, multiplied by one plus the fraction of the CFV these invalid voxels amounted to (assuming that the added layers would include an equal fraction of invalid voxels). The CFV was then defined as the part of this volume which remained after excluding the GTV voxels. Figure 4.2(a) shows an illustration in 2D of this approach. After having used this approach for different dose distributions it was deemed inappropriate, since for some positions and sizes of the CFV versus GTV, the CFV took a very irregular shape with subvolumes extending far from the centre of mass.

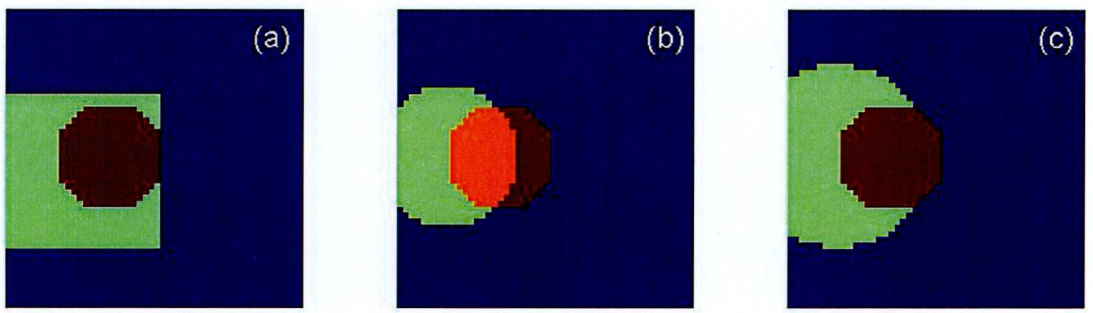


Figure 4.2: Illustration in 2D of the different methods considered for identifying the CFV (green area). The main difficulty was to make the shape of the CFV adapt to e.g. the GTV surface (red area). In (a) the CFV was a square (cube in 3D), where the size was increased to compensate for excluding GTV voxels. In (b) the FSU survival in round (sphere shaped in 3D) CFVs was calculated with a convolution method, but this could not account for adapted shapes of the CFV. The orange area represents elements shared by the CFV and the GTV. The method actually adopted, (c), identified the CFV as the valid voxels (non-GTV) closest to the CFV centre.

Instead, the shape of the CFV was chosen to be a sphere, in order to let all voxels be as close as possible to the centre of the CFV. The fastest way to calculate the FSU survival in these spheres was found to be by convolution of a binary 3D array, defining the shape of the CFV in one position, with the FSU survival array. This approach was much faster than the previous one, in which each CFV was identified separately. However, as the illustration in figure 4.2(b) shows, this method could not give the FSU survival for CFVs around organ surfaces where the shape of the CFV should be adapted. Different approximations were tried, but it was discovered that the 'critical' CFV generally was contiguous to the GTV, and was thus of a different shape than the basic sphere. Therefore it was crucial to find a consistent method for adapting the shape of the CFV to the GTV surface, and no approximation was reliable since several CFVs often had similar levels of FSU survival to the 'critical' CFV.

The approach finally adopted was to label each CFV by its centre element, and extend it to include the closest 'valid' voxels until the chosen size was reached (method described below). This approach is illustrated in figure 4.2(c). When building each CFV, every voxel of the organ array could be considered in a very large *for loop*, and the closest ones selected, but this would be extremely time consuming, so instead a vectorised method was adopted. The main problem encountered with this approach was the size of the arrays needed for large CFVs; some arrays were symmetrical 2D arrays with side length equal to the number of voxels in the organ array (125,000), minus the number of voxels in the GTV. With 8 bytes per element this variable could be of up to 125 GB, which is far too large for both Matlab and the operating system to handle (see http://www.mathworks.com/support/tech-notes/1100/1107.html). This was solved by an iteration where as many columns of the array as possible (representing different CFVs) were considered at a time. Since the FSU survival in each CFV is independent of other CFVs these calculations could be done either in series or in parallel.

Though this approach was conceptually acceptable it was very computationally expensive (up to 5 hours for each dose distribution on a single 2.33 GHz processor), and thus would not have been feasible in practice without the access to the high throughput computing Condor Pool at the University of Liverpool (www.liv.ac.uk/csd/escience/condor). A stand-alone executable was developed from the Matlab source code and submitted in parallel for different dose distributions to the pool of around 300 computers available

to Condor (Dell PCs with Intel Core 2 processors running at 2.33 GHz). Thus, since many patients could be simulated at once, the computation time was reduced from 4-5 months to around 36 hours for a set of 800 simulations.

Initially methods to reduce the computation time were searched for, especially for the case of a large CFV, since this resulted in a large 2D array of CFV columns, and was the most time consuming case. However, it was important that the CFV had the most appropriate shape around the GTV, and since in most positions a large CFV included GTV voxels, its shape could not be approximated without loss of consistency. The flowchart illustrates the steps necessary to identify each CFV with the vectorised method.

4.3.3 Implementation

The first step is to identify all CFV centres, which are all elements within the normal tissue of interest. These are given by the FSU distribution array. Once the indices of all CFV centre elements are collected the voxels in a region around each centre are identified as CFV candidate elements. For large CFVs these include all normal-tissue elements (i.e. all elements except within the GTV or outside the organ of interest). For small CFVs, on the other hand, the simulation time can be shortened by only considering elements within a subarray surrounding the CFV centre as CFV candidates. This subarray has to be large enough not to constrain the shape of the CFV even when the worst case (in terms of largest extent from the center) is considered. If no surfaces obstruct the CFV it takes the shape of a sphere, but close to surfaces such as organ and GTV boundaries, the shape of the CFV adapts. Thus the worst case is a CFV between e.g. the organ boundary and the GTV with a single layer of normal-tissue voxels. This means that the subarray of CFV candidates has to be large enough to fit all CFV elements in one 2D layer. All CFVs small enough for all its elements to fit on a 2D slice of the full 3D array are considered a 'small' CFV in this context. The rare situation of a GTV approaching a corner and forcing a small CFV to stretch predominantly in a single dimension was not taken into account as speed of calculation was prioritised.

As a result a 2D array, A, containing indices for elements in a 3D array is created. If 'N' is the number of CFV centres, and thus normal-tissue elements, and 'n' the number of candidate elements for each CFV, A is an N×n array (N×N for large CFVs). Each

column in A contains the indices of the CFV candidates corresponding to a given CFV centre. Further an array, B, of the same size as A is created, containing the indices of these CFV centres in rows, such that each row is identical. Thus for all CFV candidate elements in a column in A the corresponding CFV centre index is given by the same position in B.

Next, the subscripts corresponding to the linear index values in array A are entered into three arrays, A_x, A_y and A_z. Similarly the subscripts of the CFV centres in B are entered into three arrays, B_x, B_y and B_z. This makes it possible to calculate the distance from each CFV candidate element to the CFV centre; if C is an array of the same size as A and B the distances are given by

$$c^i = \sqrt{\left(b_x^i - a_x^i\right)^2 + \left(b_y^i - a_y^i\right)^2 + \left(b_z^i - a_z^i\right)^2},$$

where i is the same element in arrays A, B and C, and b_x^i represents element i in array B_x. To make sure no GTV elements are included in the CFV the values of the distance for all GTV elements are made the same as the maximum value in array C.

The values in each column of C are further sorted in ascending order, and the linear indices from A, corresponding to the first elements in this sorted array, are entered into a new array, D, until the CFVs are full. Thus, if a CFV has 100 elements D has 100 rows and N columns, and each column contains the indices of the CFV in a 3D array. The number of surviving FSUs for each of these indices in the 'FSU array' are collected and the total number in each column is entered into a vector, E. These values are the number of surviving FSUs in each CFV.

With the methodology described above contiguity of the CFV is not guaranteed since in some cases the normal-tissue elements closest to a CFV centre might be split into more than one contiguous object due to a GTV or an irregularly shaped organ, as illustrated in figure 4.3. Since the outcome (complication or no complication) only depends on the CFV with the lowest number of surviving FSUs it is enough to check that this CFV is contiguous. The centre of this CFV is called $CFVC_{min}$.

A binary 3D array, F, is created representing the critical CFV; the indices in the 3D array belonging to the CFV around $CFVC_{min}$ are found in the corresponding column in D. Functions from Matlab's image processing toolbox are used to identify the number

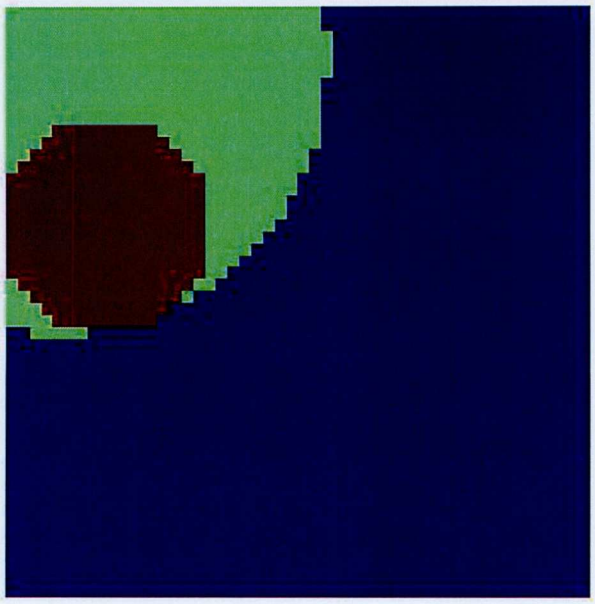


Figure 4.3: Illustration of a non-contiguous CFV (green area) divided by a GTV (red area).

of objects in F and the eventual elements belonging to smaller separate objects. For these elements the values of the distance in C are increased, similarly to how the GTV elements were penalised, and new values are chosen for this CFV into D and E subsequently. This procedure is repeated until the CFV is contiguous and then, since this CFV might not have the lowest FSU survival anymore, E is searched for the current critical CFV and its contiguity is evaluated. This procedure is repeated until the CFV with the lowest FSU survival is contiguous. It is not likely that any other CFV, currently non-contiguous, would have a lower FSU survival if contiguous, since when forcing a CFV to be contiguous elements are generally moved from close to the GTV to the periphery of the CFV where the dose is lower.

Finally the outcome is given by comparing the number of surviving FSUs in the critical CFV with the chosen value of FSU_{min} and the output of the model is given as zero or 1, for no complication or complication respectively.

4.3.4 Dose distributions

The dose distributions used in the analyses were generated separately in Matlab. Each 'patient' had a different dose distribution and the choice of characteristics for the dose distribution sets were an important part of each study design.

Dose distributions were generated using a simple beam model with the parameters randomly sampled to make each dose distribution different. Uniform, parallel beams were created with central-axis dose taken from Jordan [110] for a 6 MV beam, and a penumbra based on clinical measurements. Scatter was represented by a uniform low background dose. Figure 4.4 shows slices of three dose distributions generated to represent lung tumour treatments, using three beams.

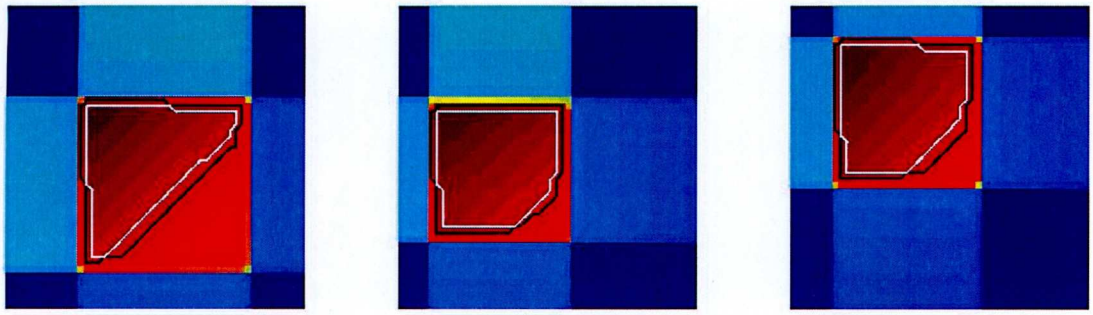


Figure 4.4: Slices of three typical dose distributions generated to represent lung tumour treatments. The PTV (black line) was defined by the isodose at 90% of the maximum dose, and the GTV to PTV margin was one voxel wide.

For organs including a target a GTV at the intersection of the beams was excluded from the surrounding normal tissue in the analysis. This was done by defining the PTV as the voxels included by the isodose at 90% of the maximum dose, and assuming a margin of one voxel between PTV and GTV, the GTV was defined as the voxels in the PTV which were not on the PTV surface; see figure 4.4. Thus the normal-tissue volume was defined by the dose distribution and the geometry of the PTV, causing further random inter-patient variation in the normal-tissue dose distribution.

Rather than attempting to closely reproduce a typical clinical setup, the purpose of this approach was to produce representative normal-tissue dose distributions. If the conventional 95% isodose had been used to define the PTV, large volumes of normal tissue would have received very high doses, due to the slow fall-off of dose of the simple three-beam arrangement chosen for lung. The inhomogeneous dose in the GTV was not a problem since only the normal-tissue dose distribution was used in the simulations. The PTV was not used for any other purpose than to define the GTV, and this procedure was chosen in order to increase the randomness and variation in the normal-tissue dose distribution, from patient to patient. Thus the small GTV to PTV margin of ~ 3 mm was of no consequence.

4.4 Discussion

The 3D model used to simulate data in this work is inevitably simplistic but attempts to incorporate the current understanding of how normal tissues/organs respond to irradiation. It has a unique potential to explore the effect of different factors on the complication probability for an organ whose response to irradiation has both a local and a global component, and it takes into account the spatial variation in dose, unlike currently used DVH-based NTCP models. Therefore, though it is not a complete model of organ response to irradiation, this model can be useful for testing the behaviour of existing NTCP models (see chapter 7). For appropriate analyses, it can generate pseudo-clinical data without limitations of patient numbers or prescribed dose ranges.

The CFV concept used in this model is similar to the critical volume tolerance suggested by Roach *et al* [70] and the variable critical volume model developed by Bonta *et al* [71], and also to some extent to the 'compartment' considered by Wolbarst *et al.* [111]. Considering an organ subregion of contiguous FSUs seems a natural choice, especially for structural radiation damage (e.g. in skin, rectum [26]) where the loss of a few contiguous FSUs might reasonably be expected to give a higher risk of complication than damaging randomly scattered FSUs, and the local 'FSU density' might be a critical parameter [112]. This is supported by the fact that the shape of the dose distribution is important for late rectal toxicity [113]. The CFV concept is also supported by the response of mouse lung to irradiation, where the tissue in sections of the lung is either largely undamaged, or completely damaged if a large enough portion of the FSUs have been destroyed by radiation [52].

Other methods of accounting for the contiguity/connectivity of damaged FSUs have been suggested by Stavreva *et al* [65] and Thames *et al* [67]. In their 'contiguous damage' model Stavreva *et al* assumed that injury to the spinal cord occurs if the number of damaged fibres exceeds a threshold, and for a fibre to be damaged a number of consecutive FSUs must be inactivated. Thus the spatial distribution of the damaged FSUs is important in their model. Similarly, the more general cluster model used by Thames *et al*, assumes that a complication occurs if the size of the largest cluster of contiguously inactivated FSUs exceeds a threshold. This model can be applied in one, two or three dimensions. In both of these models the inactivation of FSUs is

stochastic; in the contiguous damage model the probability of damaging a fibre is calculated with combinatorial analysis, whilst in the cluster model the maximum cluster size is determined with computer simulations.

Unlike these models the 3D model does not require inactivated FSUs to be contiguous (unless $FSU_{min} = 1$) to trigger a complication. Rather, it is the overall damage in the organ specific CFV which is important. The CFV is related to, but not identical with, the functional reserve. An organ with a large functional reserve has a redundancy of FSUs and retains its function even if many of them are lost. Thus, the CFV must be large and FSU_{min} low, since otherwise a complication is triggered for extensive FSU loss.

The model is also based on the FSU concept. Though the characteristics and even the existence of the FSU are controversial [114, 33], the model is only dependent on this concept via the assumption that small units of tissue need to be repopulated through local stem cell division if cells are killed by irradiation. This would still be valid for tissues which consist of several types of cells if one type is significantly more radiosensitive than the others, in which case the integrity of the FSU depends on the radiosensitivity of the most radiosensitive stem cell. The model could also be extended to include several FSU populations in one organ (e.g. parenchyma and vessels) which could respond to dose distributions in different ways (e.g. in a parallel and serial manner, respectively); for a discussion see section 1.1.3.

Since the nature of the FSU was not specified when the model was developed, it is not clear whether an injured stem cell on the edge of an FSU can be replaced by a surviving stem cell from a neighbouring FSU. The use of equation 4.1 implies that stem cells cannot migrate across FSU boundaries. Fenwick and Nahum [115, 31] presented a formula for NTCP when there are no fixed boundaries between FSUs and showed that this leads to a higher risk of local damage than if fixed boundaries are assumed. However, this formula only applies to one-dimensional FSUs. In this thesis the formulas in section 4.2 will be used.

The difference in stem cell radiosensitivity between different organs was represented by different values of α, whilst keeping α/β constant, though alternatively β could have been kept constant [116]. The effect of varying α in a population is discussed below.

One limitation of the model in its current form is that the probability of FSU inactivation depends on the local dose only, as a direct result of radiation-induced stem cell depletion. The LQ model and target cell theory do not take account of phenomena such as low-dose hypersensitivity and the bystander effect; the latter depends on signalling from surrounding damaged cells [117, 21]. Given appropriate mechanisms for these phenomena, their local or non-local effects could be readily incorporated into the model.

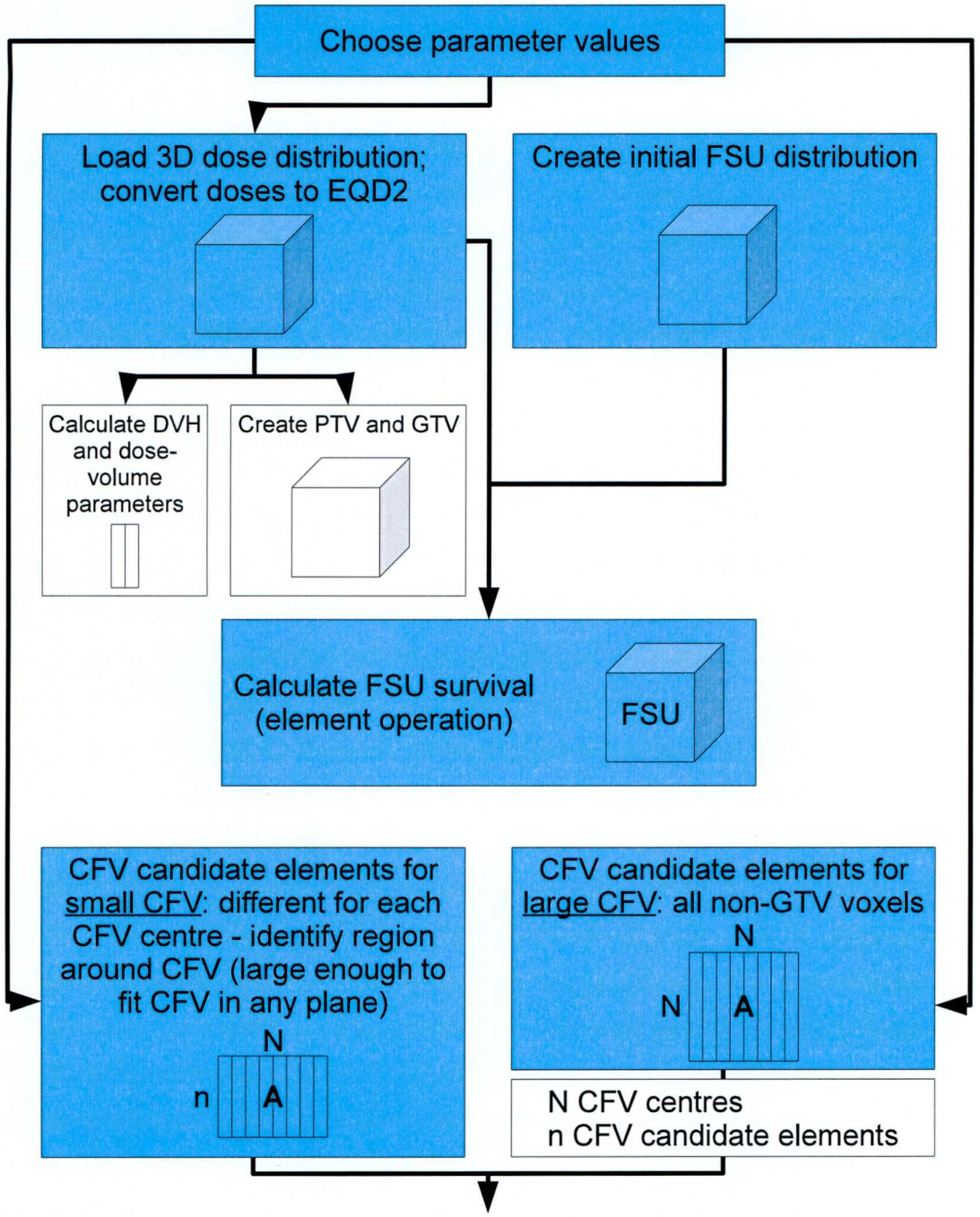
In some cases both the number of stem cells per FSU and the number of FSUs per voxel were assumed to be large, resulting in a deterministic local tissue damage function (equation (4.2)). Otherwise the number of surviving stem cells per FSU or surviving FSUs per voxel, respectively, were sampled from a binomial distribution, thus introducing a random element into the local tissue damage function. Whereas for some dose distributions, associated with a very low or a very high level of tissue damage, the overall outcome was still deterministic, for others there would then be an element of randomness in the induction of a complication for a fixed dose distribution. This is easily handled by the model unless the FSU is larger than a voxel.

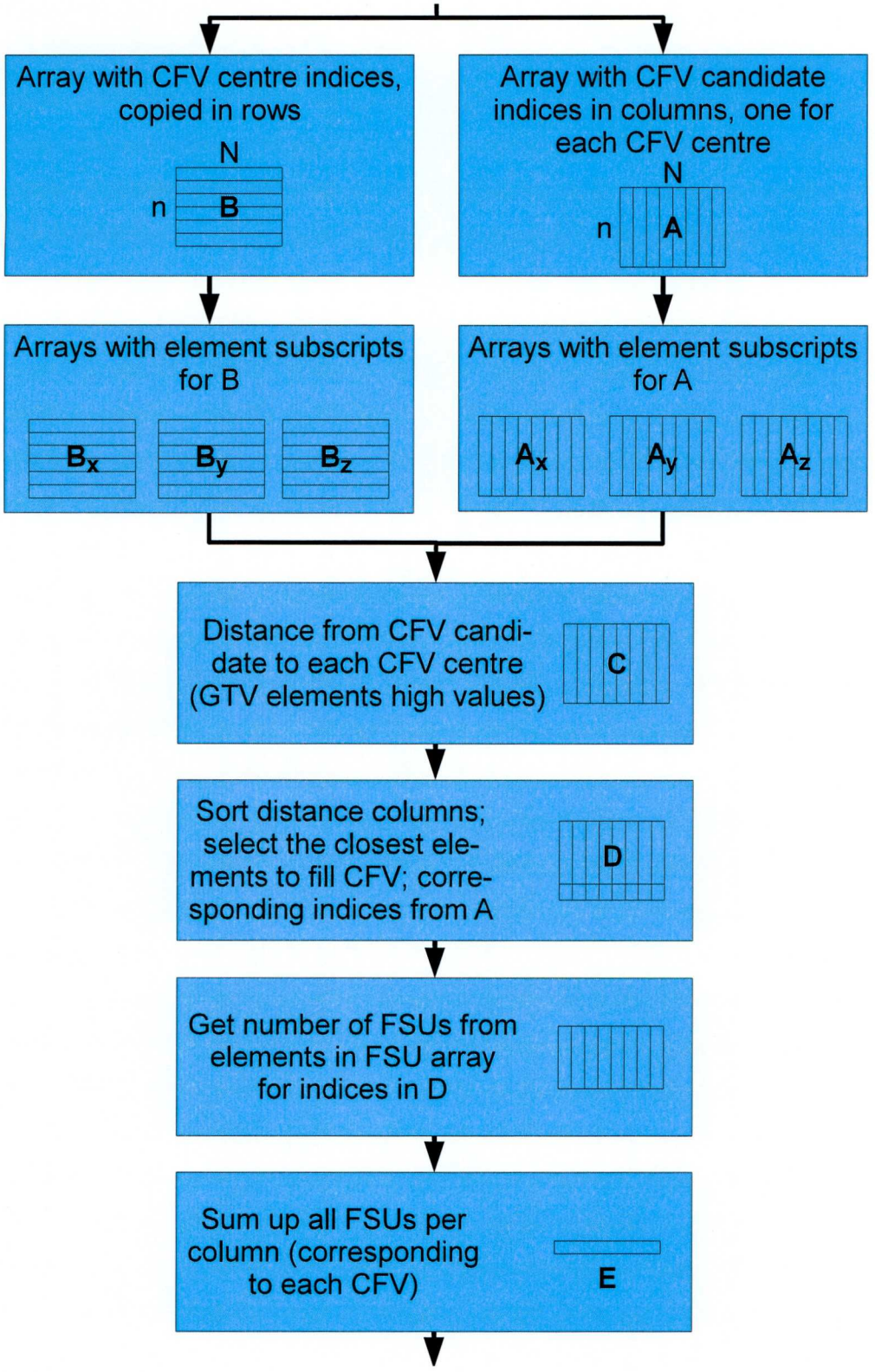
In practice, this model simulates the irradiation of an organ during radiotherapy and gives the outcome for the patient in terms of a possible complication for this organ. For a given dose distribution and FSU distribution, the result is deterministic (except when sampling N_S^{FSU}) and depends on the chosen values of the model parameters. Therefore the results correspond to what would be observed if a single patient, or a population of identical patients, could be treated with many different dose distributions. In the context of a clinical study, however, a spread in radiosensitivity as well as health status must be accounted for. The inter-patient variation in radiosensitivity and health status can be represented in the model by sampling the value of α and $\mathrm{FSU}_{\mathrm{min}}$ or N_0^{FSU}, respectively, from an appropriate probability distribution, for each patient.

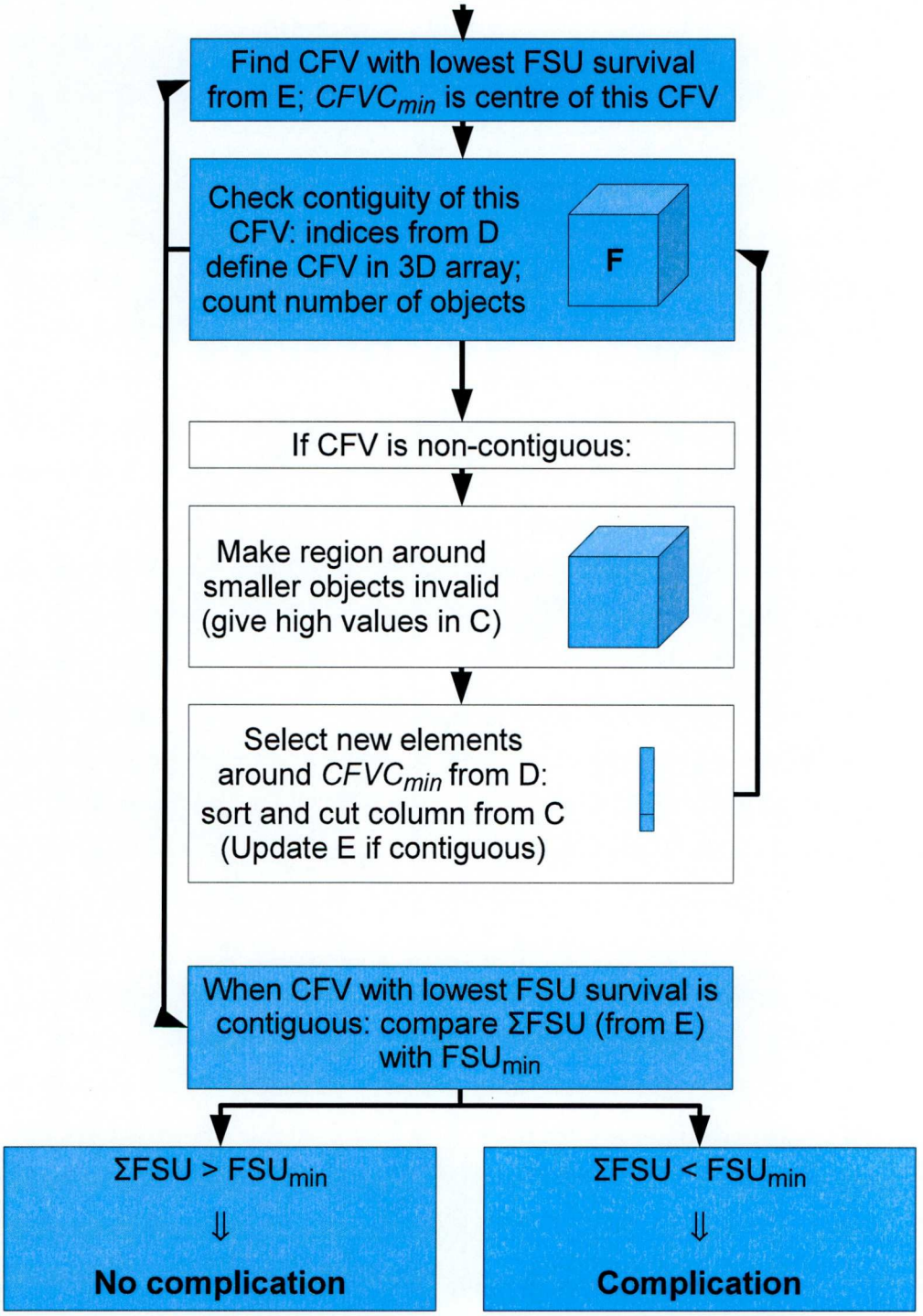
4.5 Flowchart

Here follows a flowchart illustrating the steps of the simulations described in section 4.3.

Model flowchart




Array with CFV centre indices, copied in rows
N
n
B
Array with CFV candidate indices in columns, one for each CFV centre
N
n
A
Arrays with element subscripts for B
Bx
By
Bz
Arrays with element subscripts for A
Ax
Ay
Az
Distance from CFV candidate to each CFV centre (GTV elements high values)
C
Sort distance columns; select the closest elements to fill CFV; corresponding indices from A
D
Get number of FSUs from elements in FSU array for indices in D
Sum up all FSUs per column (corresponding to each CFV)
E


Find CFV with lowest FSU survival from E; $CFVC_{min}$ is centre of this CFV
Check contiguity of this CFV: indices from D define CFV in 3D array; count number of objects
F
If CFV is non-contiguous:
Make region around smaller objects invalid (give high values in C)
Select new elements around $CFVC_{min}$ from D: sort and cut column from C (Update E if contiguous)
When CFV with lowest FSU survival is contiguous: compare ΣFSU (from E) with FSU_{min}
ΣFSU > FSU_{min}
⇓
No complication
ΣFSU < FSU_{min}
⇓
Complication

Chapter 5

The Influence of Dataset Characteristics on Correlations between Dose and Outcome

In this study simulations are carried out on the model described in chapter 4, to investigate correlations between various dose-volume parameters and the incidence of 'complications' in order to provide deeper insight into what can be expected of such analyses in clinical studies. We use the model to address a series of questions, including: How much information about an organ's dose-volume response to irradiation can be extracted from clinical studies? How does the organ architecture influence the correlation between dose-volume parameters and outcome? Why do different clinical studies often lead to different conclusions about dose-volume parameters as complication predictors [118]? This study focuses on the effect of different dose distributions on correlation analyses since a key feature of this model is its ability to evaluate the effect of full 3D dose distributions.

5.1 Dose distributions

The correlations between dose-volume parameters and outcome seem to be influenced by dose distribution characteristics [75, 68], such as variability in mean or maximum dose and volume irradiated. For an organ which encompasses the tumour the maximum dose varies little from patient to patient and the mean dose depends on the irradiated

volume. Such organs are often of parallel architecture. For an organ at risk outside the tumour, on the other hand, the maximum dose can differ greatly from patient to patient, depending on the distance to the tumour and the number of beams crossing the organ; this is often the case for serial organs. To investigate to what extent such features influence the correlations observed, the analysis was performed both for 'lung-type' and 'cord-type' dose distributions, generated as described in section 4.3.4.

For the lung-type dose distributions three beams, perpendicular to different sides of the cube representing the organ, were made to intersect at a randomly selected point. For each generated dose distribution, the weights of the three beams and the beam widths were also randomly sampled. The resulting cumulative DVHs (see figure 5.1(a)) were compared to lung DVHs of lung cancer patients treated at Clatterbridge Centre for Oncology (3D conformal radiotherapy), in order to achieve a typical range of mean and maximum dose for a given prescription dose.

Similarly, the second set of dose distributions was made to resemble the oesophagus and spinal cord dose distributions for the lung cancer patients mentioned above, without any tumour in the organ volume (see figure 5.1(b)). The dose distributions were generated as described above for the lung-type dose distributions, though in this case also the number of beams crossing the organ was sampled for each patient (average 1.7). This method of generating dose distributions ensured that the datasets used in the analysis were not limited in size or variation but still clinically relevant.

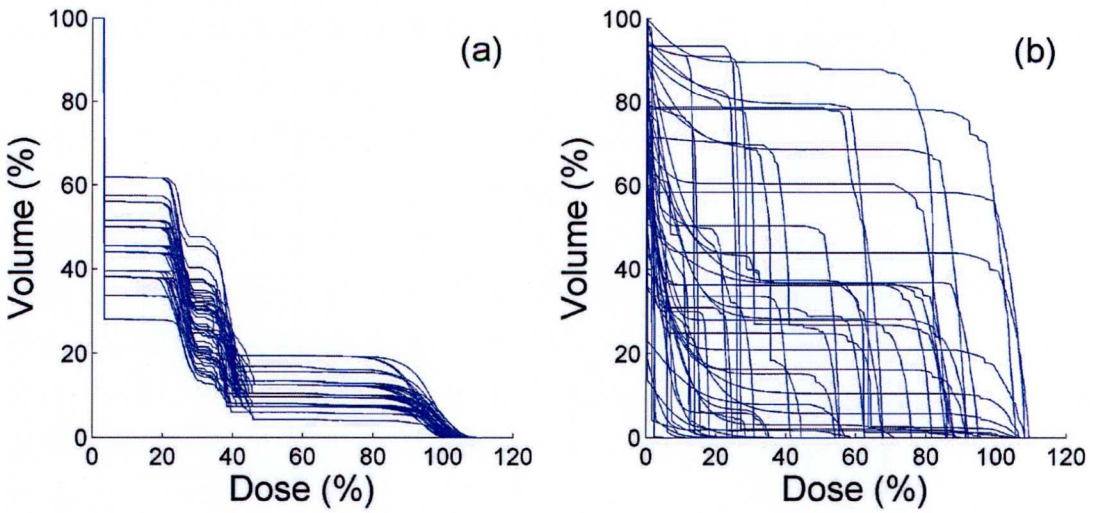


Figure 5.1: Fifty sampled dose distributions used to resemble (a) lung DVHs and (b) spinal cord and oesophagus DVHs, during lung tumour treatments.

5.2 3D model input data

The different organ types were represented by different sizes of the CFV: 0.75 for a parallel organ, 0.10 for a semi-parallel organ and 1.0×10^{-4} for a serial organ. The serial organ was represented by a very small CFV in order to guarantee typical serial-organ behaviour where the maximum dose is important. If the CFV had been a cube the side of the cube would have been 45, 23 and 2 voxels, respectively, covering the full range in the cube of $50 \times 50 \times 50$ voxels representing the organ. However, the shape of the CFV was not restricted to a cube but rather to a sphere which could adapt its shape around organ and tumour surfaces.

The value of $\mathrm{FSU}_{\mathrm{min}} = 0.56$ was kept constant; a value in the middle of its allowed range (0-1) was considered reasonable since organs seem to function despite certain tissue damage. Also, this value resulted in low numbers of complications in the clinically relevant dose range for plausible values of α and N_{sc}.

For the parallel-type organ, the value of α was chosen so that, with the lung-type dose distributions, 10% of the patient cohort with a prescription dose of 55 Gy/20 fractions experienced a complication, in line with clinical experience at Clatterbridge Centre for Oncology. For the other organ types (represented by smaller sizes of the CFV), α was chosen so that a complication rate of 50% was reached at the same prescription dose as for the parallel-type organ, in order to facilitate comparison between the organ types. These values were 0.260 Gy^{-1}, 0.137 Gy^{-1} and 0.0382 Gy^{-1} for the parallel, semi-parallel and serial organ, respectively. Although this last value might seem very low, similar values have been considered for normal tissues [119, 120, 121]. Alternatively, N_{sc} or $\mathrm{FSU}_{\mathrm{min}}$ could have been adjusted and α remained fixed for the different organ types, but N_{sc} has a much weaker influence on the response, and $\mathrm{FSU}_{\mathrm{min}}$ sometimes too strong.

The values of α listed above were also used with the cord-type dose distributions, in which case a 50% complication rate was reached at different prescription doses due to different volume effects responding differently to the different sets of dose distributions in figure 5.1.

A generic late responding normal-tissue value of $\alpha/\beta = 3$ Gy was used, and for the number of stem cells in an FSU, N_{sc}, a value of 120 was chosen based on estimates of the number of stem cells in an alveolus in the lung [122].

5.3 Simulation of clinical data

In this chapter the model described in chapter 4 was used to simulate a clinical study by calculating outcome for cohorts of patients at eight different prescription doses, and correlating calculated dose-volume parameters with the outcome using logistic regression with SPSS 16 (SPSS Inc.). This was performed for different sizes of the CFV to illustrate the cases of serial, semi-parallel and parallel tissue architectures as described above.

The number of fractions was 20 in all simulations and the set of eight prescription doses was chosen to give a range of complication rates from 1% to 50% (figure 5.2). Since the response pattern was different for each organ type, the sets of prescription doses were different and ranged from 0.75 to 5.8 Gy per fraction. This latter very high prescription dose was only necessary for the serial organ- and cord-type dose distributions because the mean maximum dose (in the organ) was 60% of the prescription dose, and because of the choice of model parameters (see section 5.2). The dose distributions were transformed into 2 Gy equivalent doses (EQD2, see section 1.1.1), using $\alpha/\beta = 3$ Gy.

The dose-volume parameters included were prescription dose, maximum dose, mean dose, V_{20} and Lyman-Kutcher-Burman (LKB) NTCP (section 2.1.1). The LKB model parameters were fitted to each organ architecture dataset separately with the maximum likelihood method (section 3.1.1). Each dataset included eight cohorts consisting of 100 synthetic dose distributions (as described in section 5.1), representing the patients in the cohort, and the outcomes were predicted by the 3D model.

5.4 Correlation analysis

A logistic regression analysis was carried out for correlations between individual dose-volume parameters and outcome, i.e. complication frequency, generated by the 3D model. The strengths of these correlations are presented in table 5.1, where the Nagelkerke R^2 parameter estimates how much of the variance in the outcome can be explained by this model fit (section 3.3.4). Thus a high R^2 value indicates a strong correlation of the dose-volume parameter of interest with the outcome.

The results in table 5.1 show that the outcome correlated well with maximum dose for organs with small CFV, as expected; for the intermediate CFV (semi-parallel organ) it correlated with mean dose, V_{20} and LKB NTCP; and for large CFV with mean dose,

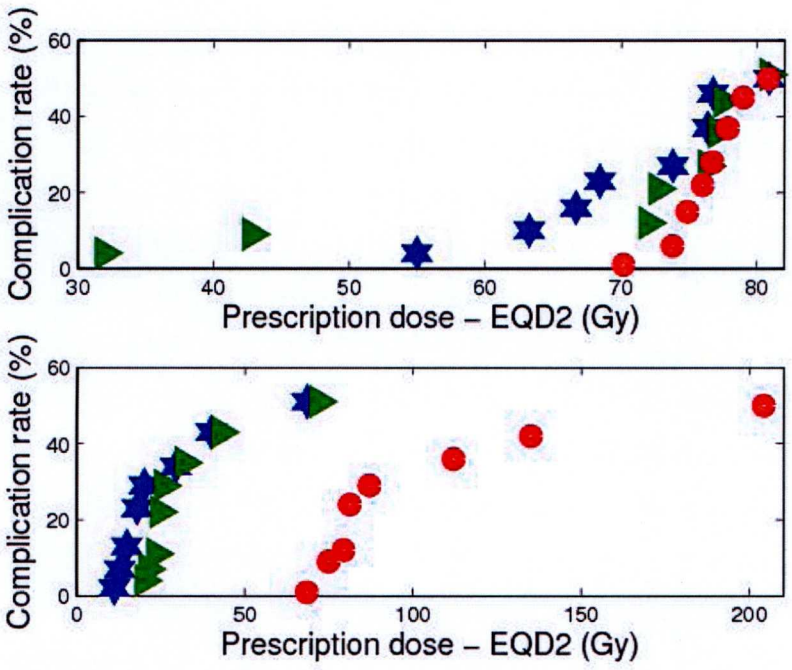


Figure 5.2: Simulated outcome data used in the correlation analysis. The stars, wedges and circles are for the parallel, semi-parallel and serial organ, respectively. (a) The radiosensitivity parameter α was adjusted to normalize the dose-response curves at 50% complication rate with lung-type dose distributions (see figure 5.1(a)). (b) Keeping the 3D model parameter values fixed, the serial organ tolerated higher prescription doses than the other organs with spinal cord-type dose distributions (see figure 5.1(b)). These curves appear to be saturating due to the large range in dose distributions in terms of maximum dose, so that some need a very much higher prescription dose than others to cause a complication.

LKB NTCP and V_{20}. This is in line with the clinical experience of maximum dose being a good predictor for spinal cord toxicity, which is often classified as a series-type complication, and mean dose and V_{20} correlating strongly with the outcome for the parallel-type complication radiation pneumonitis [118, 123]. Our results indicate that none of the considered dose-volume parameters is a perfect predictor of the outcome ($R^2 = 1$), even though no population variation in radiosensitivity or health status has been assumed here. These confounding factors can be expected to make the observed correlations weaker.

5.5 Correlations for different V_x

This analysis was also carried out for the parameter V_x (the volume receiving at least x Gy), with a range of cut-off doses x. This is often done to find the volume parameter which is the best complication predictor, and based on the result conclusions are

Table 5.1: The Nagelkerke R^2 for a logistic regression analysis for each dose-volume parameter separately, with lung-type dose distributions.

	Organ type		
Dose-volume parameter	Parallel	Semi-parallel	Serial
Prescription dose	0.18	0.15	0.22
Maximum dose	0.11	0.13	0.91
Mean dose	0.88	0.49	0.01[a]
V_{20}	0.47	0.47	0.08
LKB NTCP	0.72[b]	0.41[c]	0.33[d]

[a] $p = 0.07$; all other p-values < 0.001.
[b] $TD_{50} = 13.8$ Gy, $m = 0.089$, $n = 1.15$.
[c] $TD_{50} = 10.6$ Gy, $m = 0.25$, $n = 1.90$.
[d] $TD_{50} = 73.1$ Gy, $m = 0.047$, $n = 0.035$.

sometimes drawn concerning the volume effect and tissue architecture.

Thus far a range of prescription doses had been used since otherwise an eventual correlation between maximum dose and outcome could have been obscured because of a low 'statistical resolution'. On the other hand, in the following analysis the influence of dose distribution characteristics was clearer without the variation in prescription dose. Therefore, rather than using the data for eight prescription doses, here the 800 dose distributions were given at a single prescription dose in the simulations. The prescription dose, 3.25 Gy × 20, was chosen to give a complication rate of around 50%. Figure 5.3(a) shows how strongly V_x is correlated with the outcome for different values of x. To determine whether these correlations are due to the volume effect of the organ or to the characteristics of the dose distributions in the study, the spread in the V_x values in the dataset was studied. Figure 5.3(b) shows the standard deviation in V_x amongst the dose distributions, as a function of x.

Though the strength of correlation is not necessarily expected to follow the shape of the standard deviation curve, a small variance in a dose-volume parameter is expected to give a lower statistical resolution so that the rank order of the patients is easily disturbed by other factors influencing the observed correlations.

Typically there is a high level of cross-correlation amongst the dose-volume parameters, which can also influence the apparent correlations with outcome. Figure 5.4 shows the Spearman correlation coefficient for the range of the V_x parameters in figure 5.3, as well as for the maximum dose and mean dose since these correlate strongly with

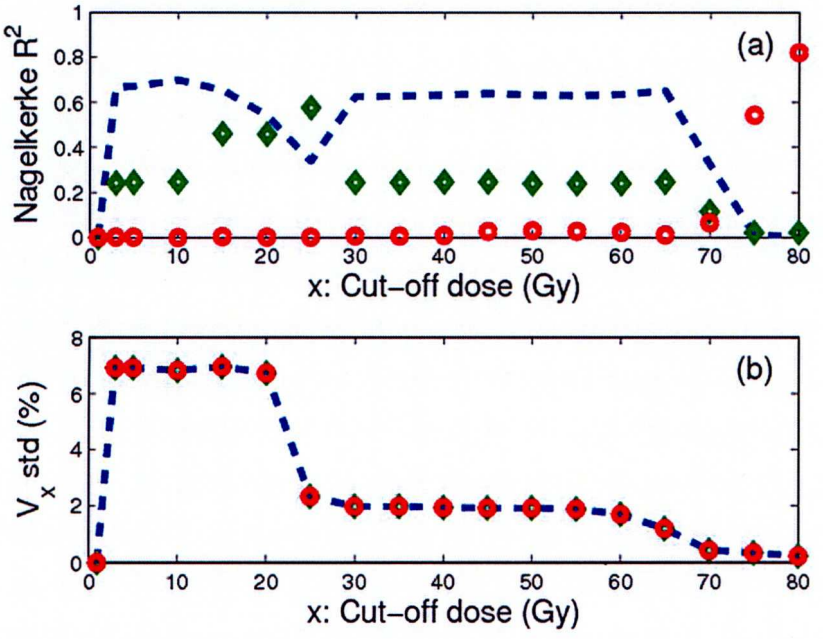


Figure 5.3: (a) Strength of correlation between V_x and outcome for different cut-off doses x. (b) The standard deviation in V_x in the set of 800 dose distributions for different x, in percent of the normal-tissue volume. The dashes, diamonds and circles are for the parallel, semi-parallel and serial organ, respectively.

outcome for serial versus parallel organs. The V_x parameters which correlate strongly with these are also expected to show strong correlations with outcome. Since the same dose distributions were used for all organ types, figure 5.3(b) and figure 5.4 were the same for all three datasets, whilst the correlation in figure 5.3(a) depended on the outcome and was different for each organ type.

A comparison of figures 5.3(b) and 5.4 with figure 5.3(a) indicates that for the parallel organ the cross-correlation of the V_x parameters with the mean dose explains the observed strengths of correlation, since the mean dose correlates strongly with outcome. Similarly, for the serial organ, the V_x parameters which correlate strongly with the maximum dose also correlate with outcome. As the outcome for the semi-parallel organ does not correlate strongly with the mean or the maximum dose, the curve shown in figure 5.3(a) cannot be explained by any cross-correlation with these parameters. Instead it largely follows the standard deviation curve shown in figure 5.3(b), except that V_{25} seems more important and V_5 and V_{10} less important. This is partly explained by a cross-correlation of V_{15}-V_{25} with D_{10}-D_{20} which were found to correlate with outcome. This issue is further discussed below. The similarity between the curves in figures 5.3(a)

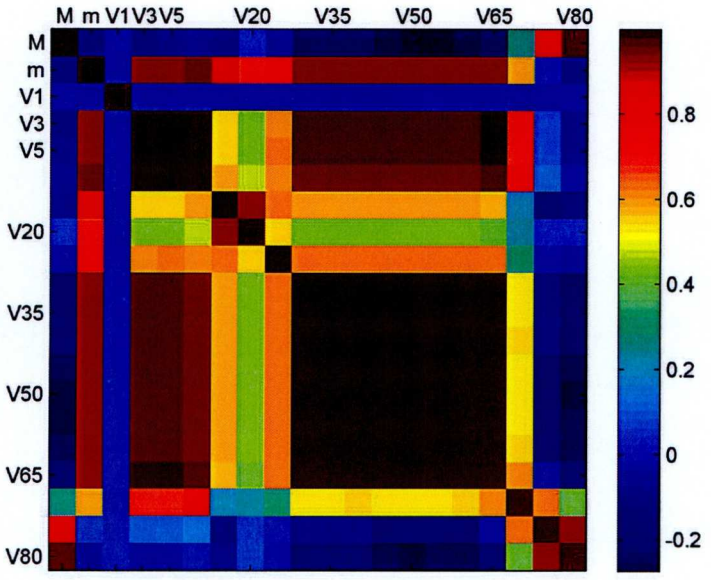


Figure 5.4: Spearman correlation coefficient for cross-correlation between dose-volume parameters (M = maximum dose, m = mean dose).

and (b) for the semi-parallel organ could mean that no V_x parameter in particular is important for this organ, so that a low standard deviation obscures the low level of correlation whilst it is boosted where the standard deviation is high.

5.6 Correlations for different D_x

Further, the correlation between D_x and outcome for different cut-off volumes x was investigated (figure 5.5(a)), as well as the corresponding spread in D_x values in the dataset for the different cut-off values x (figure 5.5(b)). D_x is the lowest dose given to the hottest $x\%$ of the organ. This is a convenient parameter since, like V_x, it can be read straight off a cumulative DVH. It can also conceptually be linked to the CFV parameter considering that the dose to the hottest contiguous volume influences the response of the 3D model. Thus $D_{0.01\%}$, D_{10} and D_{75} might be expected to be good predictors for the serial, semi-parallel and parallel organ, respectively, though not perfectly so, since the FSU kill in a CFV only relates indirectly to the minimum dose to this volume, and the D_x parameter does not necessarily correspond to a contiguous volume. Figure 5.6 shows the Spearman correlation coefficient for the range of D_x parameters in figure 5.5, as well as for the maximum dose and mean dose.

As in the case of V_x, the strong correlation of maximum dose and that of mean dose with the outcome, and their cross-correlation with the D_x parameters, explain the shape

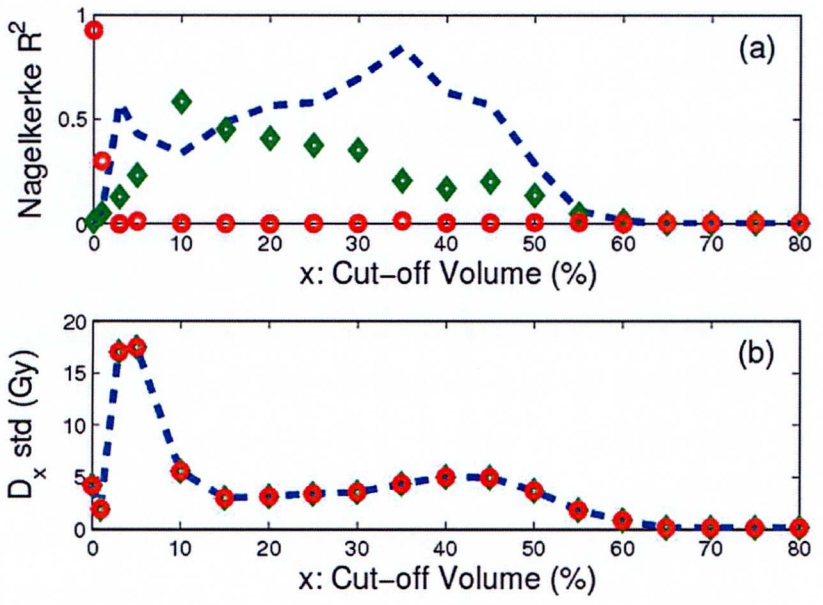


Figure 5.5: (a) Strength of correlation between D_x and outcome for different cut-off volumes x. (b) The standard deviation in D_x in the set of 800 dose distributions for different x. The dashes, diamonds and circles are for the parallel, semi-parallel and serial organ, respectively.

of the correlation curve in figure 5.5(a) for the serial and parallel organ, respectively, as is evident from a comparison with figure 5.6. The D_x parameter which correlates the most strongly with the outcome for the semi-parallel organ is D_{10}. Since there is cross-correlation between D_{10} and D_{15}-D_{30}, these also correlate rather strongly with the outcome, and around D_{45} there is a slight peak corresponding to the increase in the D_x standard deviation at D_{45}.

As expected the dose received by the hottest 10% of the organ was important for the semi-parallel organ since the CFV was 10% for these simulations. Indeed, here the D_x correlation analysis has picked up something which is due to the inherent organ function (in this case the 3D model volume effect). This also explains the correlation between V_{25} and the outcome (and the cross-correlation between V_{25} and D_{10}), as the average isodose including the hottest 10% of the organ was between 20 and 25 Gy. Similarly, $D_{0.01\%}$ is important for the serial organ (CFV $= 1.0 \times 10^{-4}$), but since the organs were not irradiated to volumes larger than 60% with any significant dose, D_{75} is not found to be important for the parallel organ (CFV $= 0.75$).

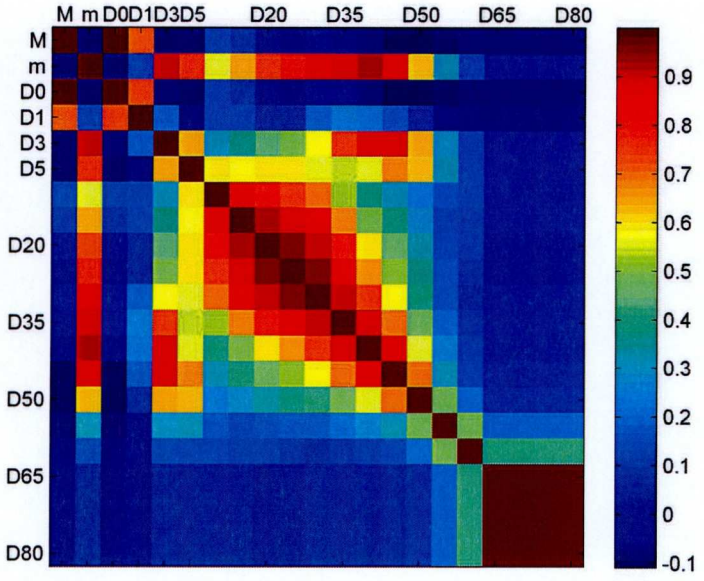


Figure 5.6: Spearman correlation coefficient for cross-correlation between dose-volume parameters (M = maximum dose, m = mean dose, $D_0 = D_{0.01\%}$).

5.7 The influence of dose-distribution characteristics

The correlation analysis was also performed for simulations with dose distributions resembling oesophagus and spinal cord dose distributions during lung cancer treatments (as described in section 5.1). Results are presented in table 5.2. In this case no GTV was extracted from the normal tissue since these organs at risk are assumed to be uninvolved by any tumour. Thus the high-dose volume is generally contiguous and regular in shape unlike the lung-type dose-distribution case. This, together with a larger variance in maximum dose, explains the stronger correlations of the outcome with maximum dose, since with similar dose distributions the dose to the CFV placed around the maximum dose (which in this case generally will be the critical CFV) will scale roughly with the maximum dose. This is not necessarily the case with an irregularly shaped high-dose region (e.g. caused by an extracted GTV) since the critical CFV will be placed around the maximum dose less consistently. Similarly the mean dose will scale with the maximum dose more consistently for uninvolved organs at risk, which explains the high correlation with mean dose for the serial organ in table 5.2.

Table 5.2: The Nagelkerke R^2 for a logistic regression analysis[a] for each dose-volume parameter separately, with spinal cord-type dose distributions.

	Organ type		
Dose-volume parameter	Parallel	Semi-parallel	Serial
Prescription dose	0.16	0.14	0.14
Maximum dose	0.35	0.63	0.98
Mean dose	0.80	0.63	0.47
V_{20}	0.46	0.77	0.12
LKB NTCP	0.72[b]	0.82[c]	0.96[d]

[a]All p-values < 0.001.
[b]$TD_{50} = 8.3$ Gy, $m = 0.32$, $n = 0.98$.
[c]$TD_{50} = 17$ Gy, $m = 0.17$, $n = 0.25$.
[d]$TD_{50} = 76$ Gy, $m = 0.027$, $n = 0.011$.

For the semi-parallel and serial organs, the correlation between the LKB NTCP prediction and outcome is much higher for the cord-type dose distributions than for the lung-type dose distributions. As shown in table 1 the LKB NTCP model, which is based on the DVHs, has difficulties in fitting to the complications generated when the high-dose region is non-contiguous, i.e. when a GTV is extracted from the normal-tissue volume. In most cases the serial organs are uninvolved organs at risk, corresponding to the cord-type dose distribution, which should make it easier to fit the LKB model to clinical serial-type complication data, if indeed the spatial distribution of FSU inactivation is important.

For parallel organs, on the other hand, the spatial distribution of FSU inactivation has a smaller effect since in this case virtually all high-dose voxels are included in every CFV. This also means that the outcome was determined by the volumes receiving the lowest doses, which varied more from patient to patient for cord-type dose distributions than for the lung-type ones where the background dose due to scatter was comparatively constant. This dependence on the low-dose volume for the lung-type dose distributions is reflected in the value of n greater than unity for parallel and semi-parallel organs.

5.8 Inter-patient variation in radiosensitivity

The above results were based on simulations using fixed parameter values, since the focus of this study was on the influence of dose-distribution characteristics on the correlations observed in a clinical study. In order to estimate the effect of an inter-patient variability

in radiosensitivity, the analysis with the lung-type dose distributions was repeated on data simulated with the value of α randomly sampled from a log-normal probability density function with a standard deviation of 25%. This large standard deviation was chosen in order to accentuate the effect on a correlation analysis.

For each organ type a thousand treatments with different dose distributions and with different values of α were simulated at each prescription dose, in order to establish confidence intervals for the dose-response curves. Figure 5.7 shows the results of the simulations, with and without variation in radiosensitivity.

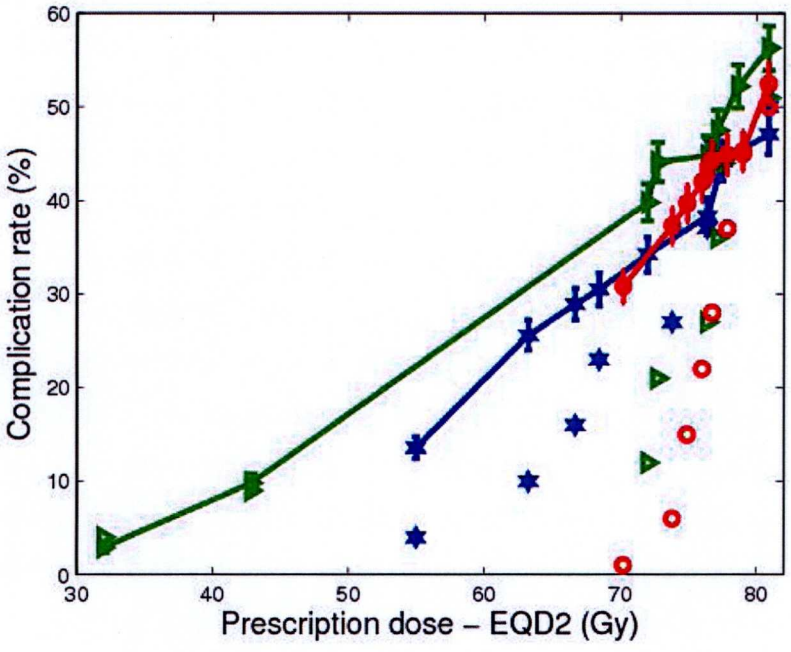


Figure 5.7: Simulated outcome data for lung-type dose distributions, with (line and errorbars) and without (only symbols) inter-patient variation in radiosensitivity. The stars, wedges and circles are for the parallel, semi-parallel and serial organ, respectively. The data points without variation in radiosensitivity are the same as in figure 5.2. The confidence intervals for the data with sampled radiosensitivity are for one binomial standard deviation based on 1000 'patients' per prescription dose, each patient with a different dose distribution.

The steepness of the dose-response curve for the serial organ was reduced considerably by introducing the variation in inter-patient radiosensitivity, as expected, and less for the parallel and semi-parallel organs. As there was a great inter-patient variability in mean dose for any prescription dose due to the different lung-type dose distributions, the slope for the parallel organ was already rather shallow and this seems to have a higher impact than the added radiosensitivity variability. The inter-patient variability in maximum dose, on the other hand, is limited and the added radiosensitivity variability has a

significant effect for the serial organ, where the maximum dose is strongly correlated with outcome. Therefore the influence of inter-patient radiosensitivity variability on the population averaged NTCP depends on the normal-tissue dose distributions as well as on the organ architecture.

This is analogous to the influence of an inter-tumour variation in radiosensitivity on the population averaged tumour control probability, which is usually large since the dose most often is uniform and constant from patient to patient. However, the effect of a variation in radiosensitivity on the correlation between a dose metric and outcome is likely to be smaller for NTCP than for TCP, especially in the case of a parallel organ, since for organs at risk there is generally a variation in response due to the variation in dose distribution from patient to patient, in contrast to the usually homogeneous dose to the tumour.

Table 5.3 shows the strengths of the correlations between dose-volume parameters and the complication rate when the inter-patient variation in radiosensitivity is taken into account. This analysis was based on 100 treatments randomly chosen from each prescription dose in order to be comparable to the values in table 5.1.

Table 5.3: The Nagelkerke R^2 for a logistic regression analysis for each dose-volume parameter separately, with lung-type dose distributions and inter-patient variation in radiosensitivity.

	Organ type		
Dose-volume parameter	Parallel	Semi-parallel	Serial
Prescription dose	0.069	0.23	0.013
Maximum dose	0.051	0.20	0.43
Mean dose	0.28	0.32	0.000[a]
V_{20}	0.18	0.23	0.001[a]
LKB NTCP	0.25[b]	0.27[c]	0.032[d]

[a] $p > 0.05$
[b] $TD_{50} = 15$ Gy, $m = 0.23$, $n = 1.0$.
[c] $TD_{50} = 19$ Gy, $m = 0.36$, $n = 0.71$.
[d] $TD_{50} = 73$ Gy, $m = 0.023$, $n = 0.040$.

The overall effect of introducing the variation in α was to weaken the correlations between dose-volume parameters and the complication rate, as expected. This had the greatest impact on the serial organ, for which even the maximum dose now only correlated weakly with the complication rate. With clinical data the maximum dose is generally found to be a good predictor of the complication rate for serial organs, which

could be due to the typically much larger spread in maximum dose for serial organs at risk, and/or a smaller natural variation in radiosensitivity than is assumed in this study.

The actual values in table 5.3 are only indicative since the standard deviation in α might be less than 25% for a clinical dataset. However, this analysis shows weakened correlations between dose-volume parameters and outcome when an inter-patient variation in radiosensitivity is introduced, and therefore the strengths of correlations observed in clinical studies are likely to be lower than those in table 5.1. This indicates that simple dose-volume parameters are inadequate as complication predictors for most organs, except the maximum dose for serial organs. As pointed out by Kong *et al* [118],'a statistically significant association or description of complication rates for populations of patients is not the same thing as a good predictor of toxicity for an individual patient'.

It is difficult to compare the strengths of correlation observed in this study with clinical dose-volume studies as generally only the p-value is reported [123].

5.9 Conclusions

A 3D model of normal-tissue response to irradiation was constructed which simulates pseudo-clinical outcome data for arbitrary dose distributions. Such simulations are ideal for studying data-analysis methodology, given the flexibility of the model and that simulations are easily performed in large numbers. Furthermore it was shown that results can be generated to resemble clinical studies, in terms of dose-volume parameters correlating with the outcome for conformal radiotherapy-like dose distributions (table 5.1).

The study of the correlations between V_x and outcome for different values of dose, x, showed that the detection of strong correlations is dependent both on the tissue architecture and to a certain extent on the characteristics of the set of dose distributions in the study, as seen for the semi-parallel organ. Moreover, if the outcome correlates strongly with maximum or mean dose, cross-correlation with this parameter will greatly influence the correlation between outcome and V_x. It is interesting that the volume, x, for which the D_x parameter correlated the strongest with outcome for the serial and semi-parallel organs, corresponded to the size of the CFV. This indicates that some information of the volume effect of an organ can be gained from a correlation analysis, if care is taken to identify other factors influencing the observed correlations.

This work may help investigators interpret the results from clinical studies and indicates how to study the influence of dataset characteristics on their results.

Chapter 6

The Radiobiological Knowledge Base for Mechanistic NTCP Modelling

6.1 Introduction

Since the mechanisms behind radiation-induced normal-tissue injury are not fully understood, predictive mechanistic NTCP modelling is not yet feasible (see section 2.3.1). However, a mechanistic NTCP model can be used as a framework for summarising the current radiobiological knowledge base, and for generating hypotheses about radiation-induced normal-tissue effects [124]. In this chapter, relevant information is collected from the literature for two important organs at risk, lung and rectum, and used to derive plausible parameter values for the 3D model described in chapter 4. It is hoped that this work will increase our understanding of normal-tissue effects, and highlight some areas that need further research before a full understanding of radiation-induced pathogenesis can lead to predictive mechanistic modelling of NTCP.

Importantly, normal-tissue effects will be considered from a mechanistic, rather than descriptive, point of view. This will be done by a reductionistic approach, where local effects on tissues (FSU inactivation), and loss of global organ function, are identified as different levels of radiation injury. The mechanisms by which the latter depends on the former vary between organs.

6.2 Pathogenesis of radiation side-effects

Radiation side-effects can be either acute (early) or chronic (late); the operational cut-off time between these phases has been set at three months after start of treatment [125]. However, in this chapter it will be more useful to distinguish between these effects based on their pathogenesis rather than on when they occur [126].

Acute effects typically result from depletion of epithelial/parenchymal cells, but are also associated with an inflammatory response [125]. For tissues with short turnover time this might happen during treatment (e.g. oesophagitis), whilst it occurs later for tissues with long turnover time (e.g. radiation pneumonitis) [126]. The ability of the tissue to recover from this damage depends on the proliferation capacity and mobility of the stem cell population.

Chronic effects, on the other hand, have a much more complex pathogenesis, including interactions between changes to the parenchyma, direct damage to the stroma and vasculature, and immunologic responses of the tissue [19]. The most common chronic effect is fibrosis, which develops months to years after radiotherapy, in most tissues. Fibrosis is predominantly a stromal lesion caused by excessive collagen deposition as a part of a wound-healing process [127, 24]. Damage to the microvasculture is also important for chronic effects due to the slow turnover of the endothelial cells, and its close connection to tissue homeostasis [19]. Loss of microvasculature also leads to direct tissue damage through ischemia [24], which can cause necrosis and ulcers. Furthermore, there appears to be a strong connection between injury to the microvasculature and radiation fibrosis [128, 13, 129].

Some organs have both an acute and a chronic phase, whilst for other organs one or the other is predominant [19]. For some organs the two phases can be difficult to separate.

6.3 Lung

The lung is a common organ at risk during thoracic irradiation and often sets the limit for local tumour control of lung cancer [130]. Since lung cancer is the cancer type with the highest mortality rate [131], it is imperative to understand how the lung responds to dose distributions, in order not to limit the dose to the tumour unnecessarily.

6.3.1 Radiation-induced lung toxicity

Irradiation of the lungs can lead to radiation pneumonitis (RP), which usually develops around 4-6 weeks after treatment [125]. This acute effect usually causes symptoms like mild fever, dyspnea and cough, and if not too severe, or if treated with steroids, resolves without chronic symptoms. During RP many type I pneumocytes, which line the alveolar surface, are lost, and the stem cells of the pulmonary epithelium, the type II pneumocytes, show an increased rate of proliferation, migrate to denuded areas and some differentiate into type I pneumocytes to replace the damaged cells. RP is also characterised by an inflammatory response [132, 10].

Chronic radiation fibrosis can develop within 1-2 years after treatment. Deposition of collagen leads to interstitial fibrosis with thickening alveolar septa, a smaller active lung volume and less efficient gas transfer. The patient might experience dyspnea, especially on exertion. Damage to pneumocytes, fibroblasts, the microvascular endothelium and the extracellular matrix are all important for the development of radiation fibrosis. However, relatively severe local damage is required to set off the fibrotic response. As long as the extracellular matrix is intact, lung tissue can be restored by type II pneumocytes, but if the internal structure is lost, tissue repair with scar formation follows, which is the initial phase of fibrosis [23, 132, 19].

In this thesis only RP will be modelled, and though this is an acute normal-tissue effect, associated with depletion of parenchymal cells, the turnover of the pneumocytes is slow enough for time effects to be negligible, and generally symptoms do not appear before the end of treatment.

The shape of the lung was not considered crucial to the results of simulations, and therefore the full array, excluding the GTV, was chosen to represent the total paired lung volume; see figure 6.1.

6.3.2 Modelling *local* tissue damage

Dose response of the FSU

On a local level, tissue damage is modelled by inactivation of FSUs in the 3D model. Each FSU responds independently to irradiation, and thus only the local dose (not the volume irradiated) influences the survival probability of an individual FSU. The

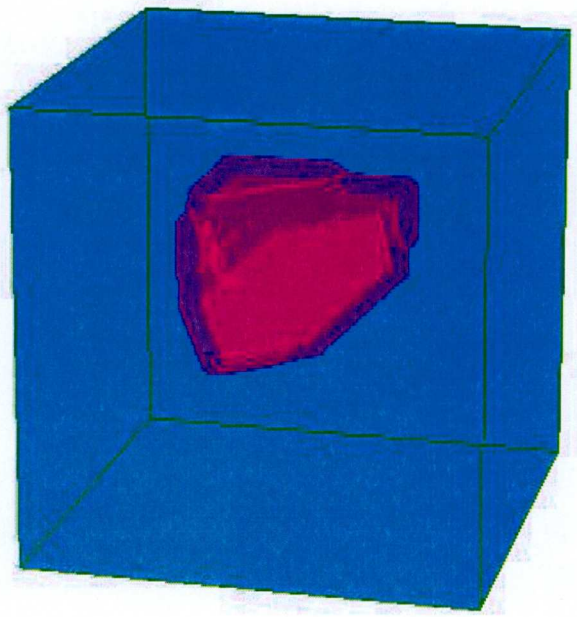


Figure 6.1: Geometrical representation of a pair of lungs in the model. The blue volume is normal lung tissue, and the purple volume is the GTV, which is not included in the normal lung.

dose-response of a tissue therefore depends on the number of stem cells per FSU (N_{sc}) and the radiosensitivity of these stem cells (α). The radiosensitivity may vary between patients, and since generally only population based data are available it is reasonable to consider the *average* radiosensitivity, $\bar{\alpha}$. If the threshold dose for FSU inactivation is known, an iso-effect relationship between $\bar{\alpha}$ and N_{sc} is given by the equation for the probability of FSU inactivation (see equations 1.2 and 4.1):

$$P_{\text{FSU}} = \mathrm{e}^{-N_{sc} \cdot \mathrm{e}^{-\alpha n d - \beta n d^2}} \tag{6.1}$$

This relationship is shown in figure 6.2, for the threshold dose estimated as 8.4 Gy (see below). Importantly, this shows the range of possible numbers of stem cells per FSU, given the plausible range for $\bar{\alpha}$; see below.

The threshold dose of 8.4 Gy was estimated based on the results of a study by Gopal *et al.* [133], where the loss of diffusion capacity in the lung was related to the local dose. Here 8.4 Gy is the lowest total dose, normalised to 2-Gy fractions ($\alpha/\beta = 3$ Gy), for which a local tissue effect was detected. Although this was a study on only 26 patients, the 95% confidence interval for the threshold dose was small (7.0-9.8). Still,

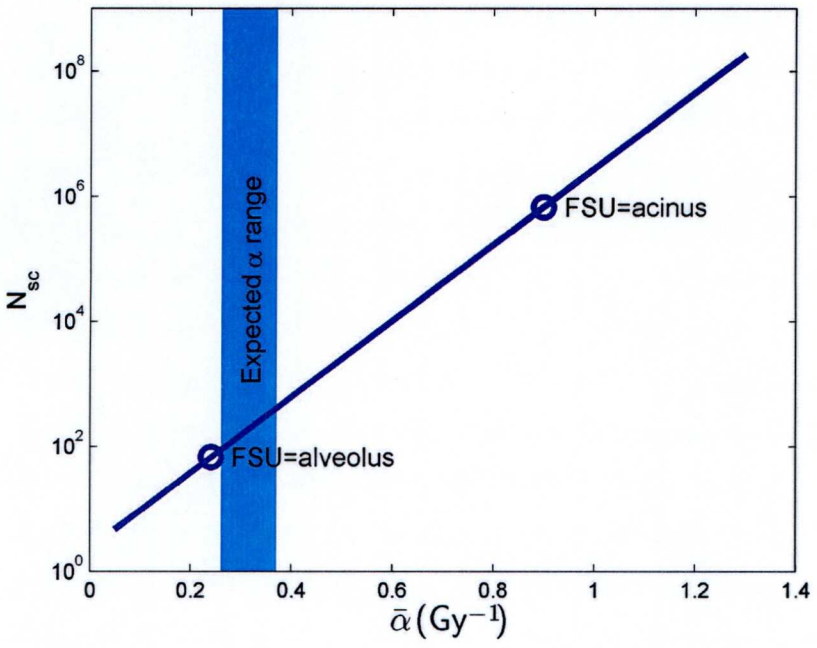


Figure 6.2: Values of $\bar{\alpha}$ and N_{sc} resulting in 90% FSU survival for 8.4 Gy. The expected range of α, and the combinations of $\bar{\alpha}$ and N_{sc} associated with the alveolus and acinus as FSU, respectively, are indicated; see section below about the FSU of radiation pneumonitis.

this value should be considered approximate since it is based on a model relating global lung function to the local dose, rather than measurements of the actual local effect. Despite this drawback the loss of diffusion capacity seems to be the most relevant measure of the local effect of radiation, since it measures the actual gas exchange in the lungs [133]. For other endpoints, studies on local lung function in terms of loss of perfusion could be useful. In contrast, these suggest a much higher dose threshold for tissue damage, with continuously increasing damage up to 60 Gy [134, 135, 136] (c.f. figure 6.3).

The fact that most of the patients in the study by Gopal *et al.* [133] also received chemotherapy is another important source of uncertainty in the chosen threshold dose. In many studies on 3D conformal radiotherapy for lung tumours the volume receiving at least 20 Gy (V_{20}) has been associated with RP [118], which suggests that the lower dose found in Gopal's study could be due to a sensitising effect of the chemotherapy. However, such findings are most likely treatment technique dependent [107, 137], and whether the lower dose threshold (V_5) indicated by recent studies, on data from IMRT for mesothelioma [138] and 3D conformal radiotherapy for oesophagus cancer [139], is a result of the

different radiotherapy treatment technique, or the influence of chemotherapy/surgery, is not clear.

Therefore, until more reliable data for the threshold dose associated with local loss of diffusion capacity is available, it will be assumed to be 8.4 Gy. However, such low doses can only cause a low level of FSU inactivation, and only contribute to the complication probability in addition to volumes at higher dose [79]. Also, since up to 13 Gy total lung irradiation is tolerated by some patients [140], the slope of the dose-response curve for the FSU cannot be too steep.

Figure 6.3 shows this dose-response for one combination of parameter values from figure 6.2: $\bar{\alpha} = 0.24$ Gy^{-1} and $N_{sc} = 67$. This combination gives 90% FSU survival at 8.4 Gy, and a slope shallow enough to allow for some FSU survival even at 13 Gy, which agrees with the experience from total lung irradiation mentioned above. The slope of this curve is also in line with the local dose-response reported by Gopal *et al.* [133] where regions of < 6.3 Gy_3, 6.3-13.4 Gy_3 and > 13.4 Gy_3 (EQD2) were associated with a 0%, 70% and 90% decrease in diffusion capacity respectively.

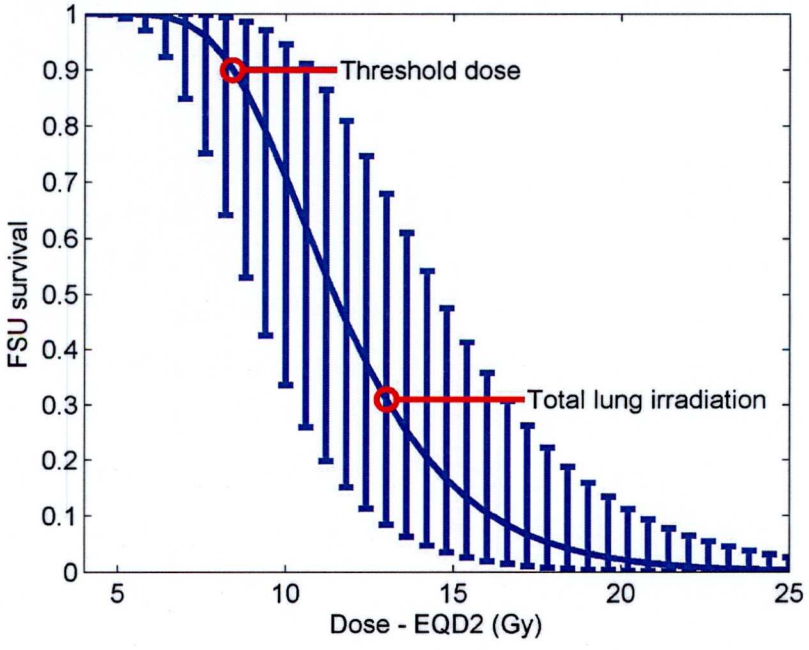


Figure 6.3: Median FSU survival with 5 and 95% quantiles for different local doses; $N_{sc} = 67$, $\bar{\alpha} = 0.24$ Gy^{-1} and $\sigma_\alpha = 15\%$. The position and slope of the curve have been guided by the threshold dose for local loss of diffusion capacity, and the dose limit for total lung irradiation.

The range of FSU survival shown for each dose in figure 6.3 is based on calculations for 1000 'patients', where the radiosensitivity for each patient is sampled from a log-normal distribution with mean 0.24 Gy^{-1} and a standard deviation of 15% (assuming a fixed $N_{sc} = 67$). Interestingly, these results indicate a large variation in dose-response of the FSU between patients, even for a moderate variation in radiosensitivity.

Furthermore, this curve displays almost complete FSU inactivation for doses above 25 Gy. This indicates that for the endpoint of radiation pneumonitis maximum local effect is reached at a relatively low dose. However, the dose effect for other endpoints, such as fibrosis, may saturate at higher doses, as indicated by the studies on perfusion mentioned above.

There is little data on organ specific normal-tissue radiosensitivity available in the literature. Two studies report values for the LQ parameters for lung, derived from radiobiological modelling [121, 141]. Very low values of α were reported (0.03-0.19 Gy^{-1}); however, these models did not include any volume effect but assumed that normal-tissue complication probability depended directly on FSU inactivation. Considering the large volume effect of the lung the radiosensitivity is expected to be greater, since a significant amount of FSUs must be inactivated before symptoms are observed. As expected, modelling including a volume effect, on mouse lung data, resulted in α values between 0.26 and 0.37 Gy^{-1} [52].

The FSU of radiation pneumonitis

In the same paper Stavrev *et al.* advocated the acinus as the FSU of the lung, since the size of this structure matched the FSU size predicted by their model, and it seems anatomically possible for cells to migrate freely throughout the acinus. However, the acinus of the human lung is a relatively large structure containing over 10,000 alveoli [142], with around 67 type II pneumocytes [143] in each; type II pneumocytes are the stem cells of the pulmonary epithelium [144, 142]. This means that the acinus contains over half a million stem cells, which according to figure 6.2 leads to an unreasonably high value of $\bar{\alpha}$.

Timmerman *et al.* [36] on the other hand suggest that stem cells cannot migrate between alveoli, and conclude that the alveolus is the FSU of the lung. In this case the value of N_{sc} would be 67 (on the average), and then a 90% FSU survival probability

at 8.4 Gy would be reached for $\bar{\alpha} = 0.24$ Gy^{-1} (figure 6.2). This value of α is close to its expected range. Another indication that the alveolus is the FSU, rather than the acinus, is given by the dose-response curve for the acinus as the FSU, which would be much steeper than the one shown in figure 6.3 (for the alveolus as the FSU), and would have no FSU survival at all at 13 Gy.

Thus, the FSU of the lung is identified as the alveolus and the approximate parameter values associated with local tissue damage are $N_{sc} = 67$ and $\bar{\alpha} = 0.24$ Gy^{-1}. The average number of alveoli in a normal pair of lungs is 500 million with standard deviation 180 million [145]. In the 125,000 voxel array of the model this means 4,000 FSUs per voxel. A lower mean FSU density of 3,600 FSUs/voxel was chosen, reflecting the frequent poor health status amongst lung cancer patients.

Local tissue damage in the lung is expected to follow the dose response function shown in figure 6.3. The importance of low doses indicated by these results is in agreement with the classification of the lung as one of the most radiosensitive tissues [125]. To be able to explain why much higher doses are frequently administered to the lung without causing clinical symptoms, the link between local tissue damage and the loss of global organ function must be investigated.

6.3.3 Modelling loss of *global* organ function

In the 3D model the loss of global organ function is determined by the critical functioning volume (CFV), as well as the distribution of surviving FSUs. A complication is triggered only if the number of FSUs falls below FSU_{min} in any one CFV. The size of the CFV is related to the functional reserve (redundancy of FSUs) and the volume effect of the organ (section 1.1.2).

The ultimate function of the lung is gas exchange, which is dependent on a sufficient number of functioning alveoli [27]. As such the CFV could be expected to encompass the total volume of the organ, and FSU_{min} might be given by the total lung volume minus the functional reserve[1]. However, there is evidence of more complex volume effect mechanisms in the lung.

With their mouse model Stavrev *et al.* [52] found that at low doses a portion of the FSUs were injured and recovered independently of each other, with the damaged FSUs

[1] FSU_{min} is the fraction of surviving FSUs each CFV needs in order to avoid a complication, whilst the functional reserve is the fraction of the organ volume which can be lost without causing a complication.

randomly distributed within the irradiated volume. In contrast, above a threshold of tissue damage (i.e. local dose) all the FSUs in a surrounding subvolume of the lung were inactivated, and therefore no surviving FSUs occurred randomly in the volume. This additional tissue damage throughout a contiguous region might be the result of an inflammatory process, triggered by the initial radiation-induced tissue damage. Thus, the radiation-induced tissue damage in itself might not have caused significant loss of organ function, but since it also triggers inflammation-induced tissue damage a complication occurred. In this case the CFV represents the large volume over which this inflammation-induced damage spreads, after the radiation-induced damage has caused the number of functioning FSUs in the volume to fall below FSU_{min}.

These volumes are illustrated in figure 6.4, where for simplicity partial uniform irradiation is assumed (only A irradiated); A is the volume irradiated and damaged, B is the volume damaged indirectly (e.g. through inflammation) by the irradiation of volume A, and C is the (unirradiated and) undamaged volume. The CFV corresponds to the volume A+B, and FSU_{min} to B.

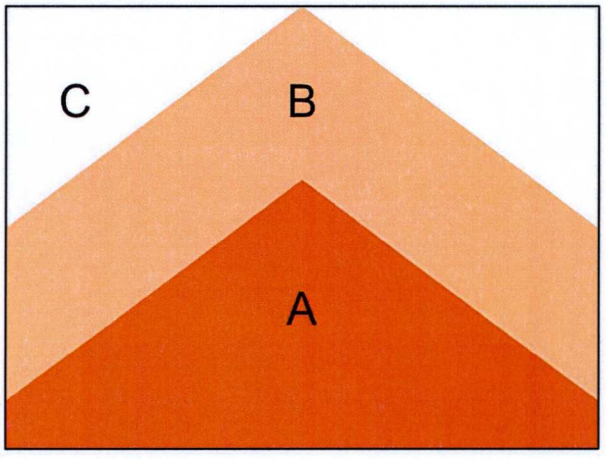


Figure 6.4: Illustration of the application of the CFV concept to the lung. A = irradiated volume, B = unirradiated but damaged volume, C = unirradiated and undamaged volume.

Based on the size of the smallest damaged volume causing a complication in the mouse [52], the size of the CFV was approximated to be 75% of the total lung volume. Simulations with the model were then carried out for different values of FSU_{min}. One hundred lung-type dose distributions were delivered at a prescription dose resulting in an average mean lung dose of 29 Gy, which corresponds to the EUD expected to give 50% complication probability, based on the LKB model fitted to clinical data [93]. This

prescription dose was 7 Gy × 20 for the dose distributions of the dataset. Simulations were repeated for different values of FSU_{min}, until 50% complication rate was achieved, between 0.46 and 0.47.

The plausibility of this value is confirmed by the volume which must be spared from radiation damage, in order to avoid pulmonary complications, suggested by a correlation analysis for a threshold dose of 5 Gy [75]. The value of around 60% of the lung volume[2] from this study corresponds[3] to $FSU_{min} = 0.47$. However, the patients in this study were treated with chemoradiotherapy, and a much lower value of around 0.11 is suggested by a large study by Seppenwoolde *et al.* [43], where it was found that 33% of the total lung volume must be spared from doses above 13 Gy.

Thus the clinical evidence for the value of FSU_{min} is conflicting, and of limited value for this analysis, since it is influenced by unknown dataset characteristics. The value of 0.47, resulting in the appropriate complication rate mentioned above, will be used.

6.3.4 Inter-patient variation in radiosensitivity

When constant parameter values are used all patients are assumed to be identical, except for the dose distribution (which depends on tumour position and size). To represent the inter-patient variation in radiosensitivity, α was sampled from a log-normal distribution, since α is bounded by zero, with a standard deviation of $\sigma_\alpha = 25\%$. This would be expected to flatten the dose-response curve, if the complication rate is plotted against different prescription doses (simulating a dose escalation study). However, this variation in α had little or no impact on the complication probability when a set of lung-type dose distributions was considered, since the range of complication probabilities associated with the different dose distributions was large enough to overshadow the effect of the variation in radiosensitivity; see discussion in section 5.8.

Therefore, the value of σ_α cannot be deduced from such a curve in a dose escalation study. Instead clues can be obtained from the expected range in the local dose response curve of the FSUs (figure 6.3), and from the variation in radiosensitivity between different mouse strains, which suggests $\sigma_\alpha = 18\%$ [52]. To guarantee adequate FSU

[2]Only an absolute volume was given, but a relative value was calculated assuming normal lung volumes.

[3]When deriving FSU_{min} from the volume which must be spared, it is assumed that the volume outside the 'critical' CFV receives doses below the threshold, and thus FSU_{min} = (volume spared - (1-CFV))/CFV.

survival at doses tolerated for total lung irradiation (see discussion above), a similar value of $\sigma_\alpha = 15\%$ will be used for lung simulations.

6.3.5 Inter-patient variation in health status

In a clinical setting patients will differ, not only in stem cell radiosensitivity, but also in health status and lung volume, and therefore in functional reserve. This is represented in the model by an inter-patient variation in pre-treatment (uniform) FSU density (i.e. the number of pre-treatment FSUs per voxel, N_0^{FSU}). This is suitable since previous lung damage (caused e.g. by smoking) causes damage throughout the lung tissue.

Perhaps more importantly, it has been shown that the clinical manifestation of radiation pneumonitis is correlated to the strength of the inflammatory response to irradiation, and that this is different for different individuals [146]. It would be more intuitive to represent this by a varying $\mathrm{FSU}_{\mathrm{min}}$, but in practice the effects of varying N_0^{FSU} and $\mathrm{FSU}_{\mathrm{min}}$ are equivalent for a uniform FSU distribution, since both affect the ultimate comparison of the number of surviving FSUs in the 'critical' CFV to $\mathrm{FSU}_{\mathrm{min}}$, which decides whether a complication occurs.

One source of information concerning the variability in health status among lung cancer patients is the level of correlation observed between mean lung dose (MLD) and complication probability. This variability acts as a confounding factor in dose-volume analyses, so that increasing variability in health status should result in weaker correlation between MLD and outcome; see chapter 7.

This relationship was explored by simulating 1000 lung-type radiotherapy treatments with the 3D model. As expected the outcome correlated strongly with MLD, but with increasing variation in pre-treatment FSUs per voxel, N_0^{FSU}, this correlation was weakened. In these simulations N_0^{FSU} was randomly sampled from a log-lognormal distribution since it is bounded by zero and one.

Figure 6.5 illustrates the influence of this increase in the population standard deviation of N_0^{FSU}, σ_{FSU}, representing the health status variability. For each σ_{FSU} the same treatments (with the same set of 1000 dose distributions) were simulated 10 times, each time sampling new values of N_0^{FSU}. In all simulations σ_α was kept at 15%. Figure 6.5 shows the mean and standard deviation of the strength of correlation between MLD and outcome for these simulations, in terms of the area under the ROC curve. A

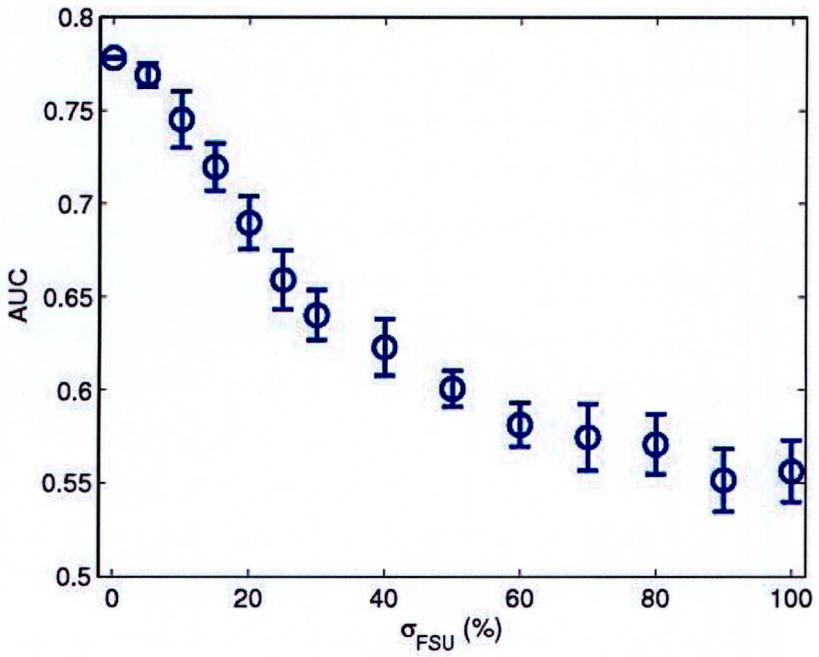


Figure 6.5: Simulations with different population standard deviations in the pre-treatment number of FSUs/voxel. The data points show the mean area under the ROC curve (AUC) for MLD, and standard deviations, for 10 simulations. The correlation between MLD and outcome decreases when the variation in pre-treatment 'FSU density' (σ_{FSU}) is increased. This corresponds to an increased level of confounding factors which weakens the link between the dose distribution and outcome. The choice of σ_{FSU} must give a realistic value of AUC.

few clinical studies have reported the AUC for MLD as a predictor of radiation-induced lung toxicity. However, σ_{FSU} cannot be determined from these reports since AUC ranges from 0.47 to 0.94 [147, 148, 149, 150, 151]. As shown in chapter 7, this variation might be due to different levels of confounding factors and complication rates, as well as different variability in MLD, in the different datasets.

Instead, a different approach was tried, where the p-value of simple univariate correlation analyses between MLD and pulmonary toxicity was reproduced. The results from three dose-volume studies on non-small cell lung cancer patients were imitated in terms of sample size and number of complications [89, 152, 153]. By simulating these studies using different values of σ_{FSU}, the p-value for a univariate logistic regression analysis of MLD and outcome could be compared to the results of these studies, and a suitable value of σ_{FSU} could be selected. Lung-type dose distributions were used, and parameter values as reported above. Each study was simulated fifty times, selecting the appropriate number of dose distributions from the set used in the simulations in the ROC analysis above, and the value of N_0^{FSU} was sampled for each patient as described

above. Table 6.1 shows the range of p-values which resulted for each of the scenarios, for one value of σ_{FSU}.

Table 6.1: The median and range (1^{st} and 3^{rd} quartiles) of the p-value for the correlation between MLD and outcome for 50 simulations, imitating the results of three clinical studies whilst assuming $\sigma_{FSU} = 25\%$.

Imitated study	median p	1^{st} quartile	3^{rd} quartile
Yorke *et al.* [153], $p = 0.04$[a]	0.18	0.027	0.37
Graham *et al.* [89], $p = 0.10$[b]	0.026	0.0019	0.12
Hernando *et al.* [152], $p = 0.005$[c]	0.0026	2.8×10^{-4}	0.021

[a] 9/49 (18%) $\geq$ grade 3 lung toxicity
[b] 22/99 (22%) $\geq$ grade 2 radiation pneumonitis
[c] 39/201 (19%) $\geq$ grade 1 pulmonary toxicity

As expected, the choice of σ_{FSU} influenced the p-value of the univariate correlation analysis, and as is evident in table 6.1 the range of p-values depended on the size of the dataset. The chosen value of $\sigma_{FSU} = 25\%$ gave representative strengths of correlation for the three scenarios; p-values both higher and lower than those observed in the three studies were generated by the simulations. This value of σ_{FSU} also led to a plausible value for the strength of the correlation between MLD and outcome of AUC = 0.66 in figure 6.5.

As seen in table 6.1, the variation in observed strength of correlation depended on the number of patients in the simulation (increasing towards the bottom of the table). Interestingly, simulations of the smallest study by Yorke *et al.*, which resulted in a low p-value, suggest that in most cases such a small study would not show a significant correlation between MLD and outcome. Thus, large patient groups are required in order to be sure to detect correlations between dose-volume parameters and outcome, when there is influence from confounding factors.

6.4 Rectum

Prostate cancer is the most common type of cancer amongst men [131]. During radiotherapy the whole gland is generally considered a target, which exposes the rectum to high doses, since the anterior wall is adjacent to the prostate. The position of the rectum makes it less mobile than the rest of the intestine, which results in the same section being consistently irradiated throughout the treatment, making the rectal wall susceptible to injury [154].

6.4.1 Radiation-induced rectal toxicity

Acute complications of the bowel occur during or soon after treatment and consist of diarrhoea and pain, mucous discharge, tenesmus and bleeding. These symptoms generally subside within 2-3 months [155]. However, some patients experience chronic rectal complications which often appear months to years after therapy [156, 157]. Rectal bleeding, faecal urgency, incontinence and more frequent bowel habits are common amongst these patients [158]. It is believed that the target cell for acute effects is the stem cell of the epithelium, whilst chronic effects are predominantly dependent on the integrity of the endothelial cells in the microvasculature [128, 24, 13, 129].

One of the most common late complications after irradiation of the rectum is bleeding [159, 160]. Whilst this is generally transient and well tolerated by patients [159, 160], severe bleeding is often persistent and requires difficult interventions [155]. The main cause of rectal bleeding is telangiectasia [161, 158, 159, 160], which appears as dilated and abnormally structured mucosal microvasculature [162, 163]. Although the pathogenesis of telangiectasia is not fully understood, it is generally attributed to damage of the microvascular endothelium [162, 163, 24, 164, 165, 13, 19].

However, telangiectasia does not explain all cases of bleeding, and not all patients with telangiectasia experience rectal bleeding [161, 160]. Another cause of rectal bleeding is ulcers, but like telangiectasia the ulcers are ultimately also due to radiation-induced endothelial cell degeneration since this leads to loss of microvasculature, and the ischemia which ensues can set off this ulcerative process [24, 165]. Thus, rectal bleeding is related to damaged endothelial cells, whether directly through telangiectasia or indirectly through mucosal ulcers, and it will be assumed that the target cell for rectal bleeding is the endothelial cell in the mucosal microvasculature.

6.4.2 Geometrical representation

In the 3D model the rectum was represented by a single voxel tube with a square cross-section; see figure 6.6. The side of the square was 2.8 cm, calculated from the cross-sectional area given by the mean rectal diameter of 3.2 cm [166], and the length of the rectum was equal to the 50 voxel side length of the array, corresponding to 12 cm. Further, the rectal wall thickness was assumed to be 2.4 mm, since this was the width of the voxel (12 cm/50 voxels), and it is close to the average thickness in normal

anatomy [167]. As shown in section 6.4.3, there are 64 FSUs/mm^2 rectal wall, which means 370 FSUs/voxel.

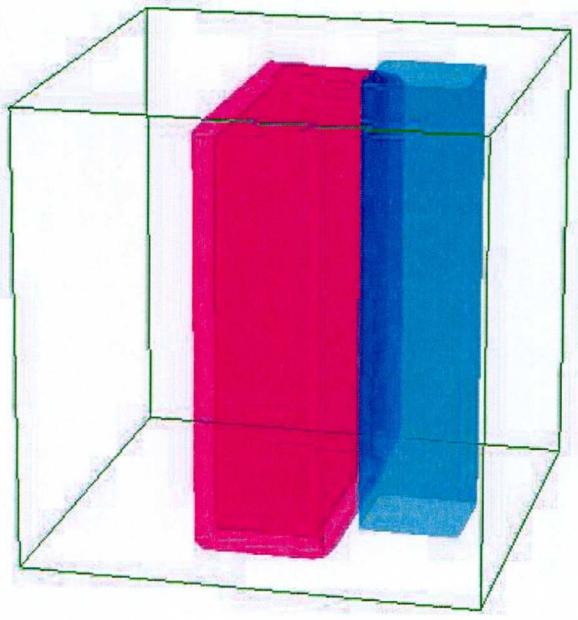


Figure 6.6: Geometrical representation of the rectum in the model. The blue volume represents the rectal wall, and the purple volume is the GTV, which does not overlap with the organ at risk. The surrounding white volume represents all other tissues and organs in the pelvis.

6.4.3 Modelling *local* tissue damage

The FSU of rectal bleeding

As for the lung above, information about the identity of the FSU of the rectal microvasculature was searched for in the literature, in order to determine values for $\bar{\alpha}$ and N_{sc}. In the rectal wall the capillaries are organised in clusters around mucosal glands which have a diameter in the order of 0.15 mm [168]. Since endothelial cells can migrate over distances of 0.3-2 mm [169], and all capillaries in a cluster seem to drain into a single venule [168], and thus are in physical connection, it is reasonable to assume that a single proliferating cell can regenerate all endothelial cells in such a capillary cluster. Thus, the FSU of the rectal microvasculature is assumed to be the capillary cluster, with a density of 64 FSUs/mm^2 smooth rectal wall, as counted from figure 2a in reference [168].

Figure 6.7 shows a schematic illustration of a segment of the rectal microvasculature. The endothelial cell is 25-50 μm long and 10-15 μm wide, and the capillaries are 7-

10 μm in diameter [170]. Kessel [170] also reported that the maximum distance to a capillary from any cell in the body is 50 μm, which is in agreement with the average inter-capillary distance of 107.2 μm in the large intestine found by Fait *et al.* [168]. Given these dimensions, around 441 capillary segments of length 2.4 mm fit in one voxel, each with a wall area of 64,000 μm^2. From the size of the endothelial cell, the number of cells per capillary segment was estimated to be 174, which amounts to 208 cells per FSU. According to Rubin [13] one in 10-20 of the endothelial cells is proliferating, which would suggest an N_{sc} of 21-42.

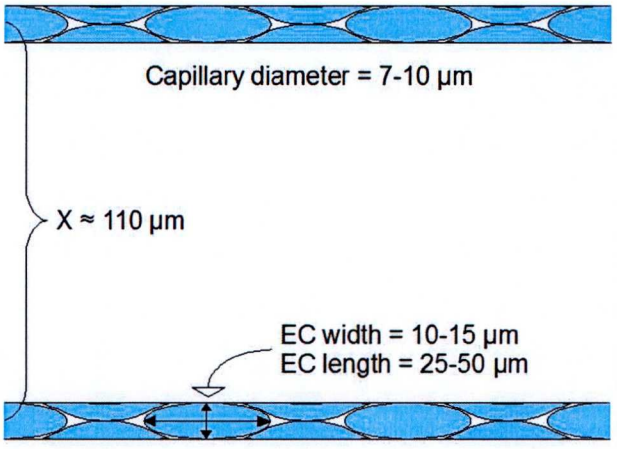


Figure 6.7: Schematic illustration of the proportions of capillary size, endothelial cell (EC) size and inter-capillary distance (X).

Unlike the epithelial cells in e.g. lung and intestine, endothelial cells retain the capacity to proliferate even in their differentiated form [171, 172]. Following cell loss, remaining endothelial cells stretch and migrate to immediately repair the endothelium, and eventually the cell density is recovered through proliferation of a small portion of the population [173, 174, 171].

The endothelial cells are also different in that they undergo apoptosis rather than mitotic cell death, if damaged by radiation. The endothelial population maintains a balance between proliferating and quiescent cells, and cells escape this cycle through apoptosis, if damaged, or through senescence (ageing), at the end of their life span or triggered prematurely by damage to the DNA. The proliferating cells are distributed randomly throughout the endothelium [171], and go through a maximum of 20-50 divisions [175]. Endothelial cell loss, caused by e.g. radiation, is compensated for by quiescent cells entering the mitotic cycle [164]. There is also evidence that stem cells from the systemic circulation can be a source of new endothelial cells [172].

Dose response of the FSU

A threshold dose for FSU inactivation of 35 Gy is assumed, based on a correlation analysis on a large dataset with varying treatment techniques and prescription doses [113], and a study estimating the local dose response to outcome with a cluster model [68]. As mentioned in section 6.3.2, the threshold dose gives the iso-effect relationship between $\bar{\alpha}$ and N_{sc}; see figure 6.8. The figure suggests that for the expected values of N_{sc} between 21-42, $\bar{\alpha}$ is in the range of 0.033-0.047 Gy^{-1}. This is in line with the low value of $\alpha = 0.01$ Gy^{-1} reported for the rat rectum [176], and 0.04-0.11 Gy^{-1} for telangiectasia [177].

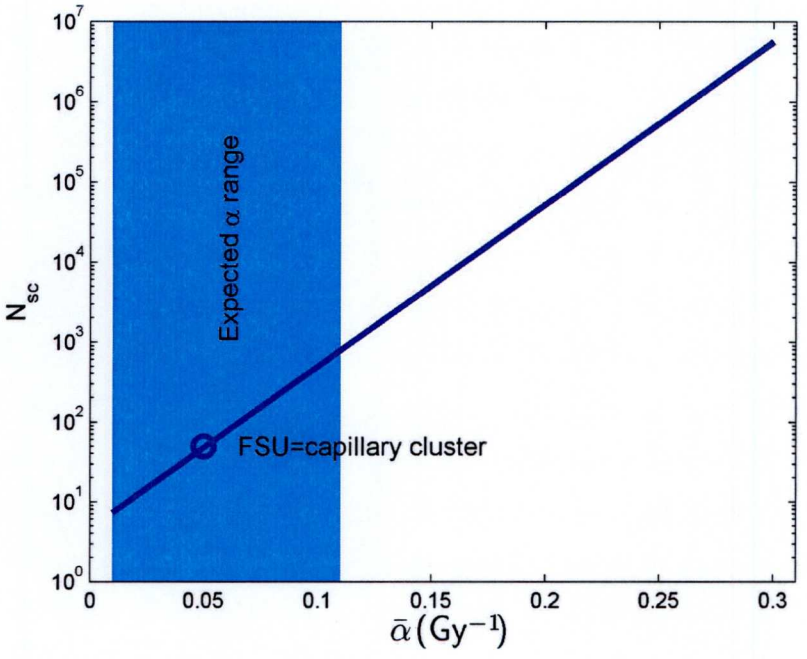


Figure 6.8: Values of $\bar{\alpha}$ and N_{sc} resulting in 99% FSU survival at 35 Gy. The expected range of α, and the combination of $\bar{\alpha}$ and N_{sc} associated with the capillary cluster as the FSU of rectal bleeding, are indicated.

Low values of N_{sc} result in very shallow dose-response curves for the FSU. The slightly higher value of $N_{sc} = 50$ in combination with $\bar{\alpha} = 0.05$ Gy^{-1} is shown to give an appropriate slope in figure 6.9, with a dose threshold of FSU inactivation around 35 Gy, major FSU inactivation at 60 Gy (60% FSU survival), and a saturation of effect above clinically relevant doses.

These results are consistent with doses ≤ 25 Gy (EQD2) considered low, ≥ 60 Gy high, and doses in between intermediate, by van Lin *et al.* [161] and Wachter *et al.* [178]. These reports might contain the most relevant dose-response data since they link

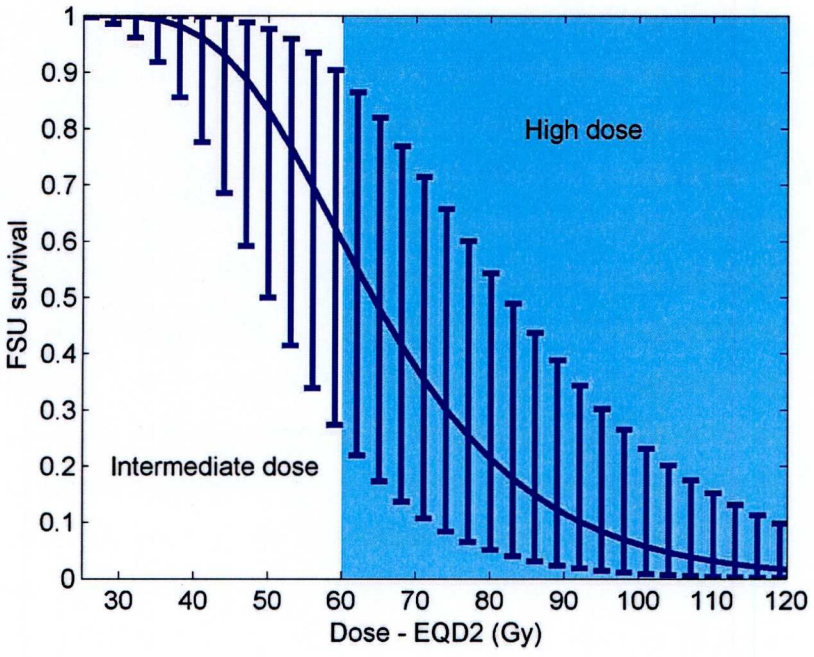


Figure 6.9: Median of FSU survival with 5 and 95% quantiles for different local doses; $N_{sc} = 50$, $\bar{\alpha} = 0.05$, $\sigma_{\alpha} = 15\%$. The position and the slope of the curve are a result of these parameter values, which follow from identifying the capillary cluster as the FSU for late rectal bleeding. Intermediate doses cause a moderate level of FSU inactivation, and high doses severe FSU inactivation.

local dose to grade of telangiectasia. They show that local doses of at least ~ 30 Gy occasionally result in low grade telangiectasia, ~ 50 Gy commonly in low or high grade telangiectasia, and ~ 60 Gy frequently in persistent high grade telangiectasia. These results compare well with figure 6.9 where the curve is the steepest at 50-80 Gy, so that these doses result in very different levels of FSU kill for patients with different intrinsic radiosensitivity.

The shallow slope of the curve in figure 6.9 also provides an explanation for the difficulty of identifying a single important dose constraint for late rectal bleeding; a range of clinically relevant doses cause an increasing level of local tissue damage, and the importance of these doses depend on the volumes treated. However, this interpretation assumes a significant volume effect for rectal bleeding. This will be discussed in section 6.4.4.

Note that here N_{sc} represents the number of *proliferating endothelial cells* in an FSU rather than a stem cell population as such. Since the putative endothelial stem cells circulate throughout the body it is unlikely that this population is influenced by local irradiation, which justifies not taking these into account for N_{sc}. This source of

new endothelial cells would rather be expected to reduce the effective radiosensitivity of the endothelium, which might explain the low values of $\bar{\alpha}$ found in this analysis. In this case a lower tolerance dose or a steeper dose response would be expected for endothelial cells in vitro compared to in vivo, since in vitro there would be no access to stem cells from the systemic circulation; and indeed endothelial cells appear to be more radiosensitive in vitro compared to in vivo [179].

The α/β ratio

As a late effect, rectal bleeding would be expected to have a low α/β ratio. However, there are several reports of values between 4.4 and 6.65 Gy for late rectal injury in general [180, 181, 176, 158], and for endothelial cell kill in particular [182]. The former could be attributed to its dependence on the endothelial integrity, or if the α/β for endothelial cells is low after all [177], to a contribution from a consequential acute effect in the rectal epithelium [181, 158]. Whatever the mechanisms though, for $\alpha/\beta \sim 6$ Gy the dose threshold for significant endothelial cell loss between 14 and 15 Gy in a single fraction [38] is consistent with the doses of clinical importance for rectal bleeding described above. Therefore $\alpha/\beta = 6$ Gy has been used in this study.

6.4.4 Modelling loss of *global* organ function

The volume effect of the rectum, with regard to late bleeding, is controversial [183]; on one hand there is a relatively strong consensus of the rectum behaving like a serial organ, and on the other hand many recommended dose constraints are for intermediate doses applied to 35-40% of the rectum (average volume from table 1 in reference [183]). The risk of rectal bleeding seems to depend *mostly* on the high-dose region, but large volumes of intermediate doses also add to the risk [183]. This gives support for the shallow dose-response of the FSU (figure 6.9), and suggests a rather large CFV for late rectal bleeding.

In a case-control study Tucker *et al.* [68] compared the size of clusters of damaged area elements in the rectal wall for pairs of patients with similar DVHs but different outcome. No fixed cluster size threshold was identified and the authors suggest that the probability of rectal bleeding increases over a range of cluster sizes. However, it was shown that there were both cases and controls for clusters of 42-72 cm^2, which

corresponds to roughly 30-55% of the rectal wall. Also, Buettner *et al.* [113] found that rectal bleeding was associated with at least 37.4% of the rectal wall receiving ≥ 48 Gy (in EQD2), and at least 59.1% of the rectal circumference receiving ≥ 59 Gy. Based on these data the CFV is estimated to be 45%.

Simulations with the parameter values derived so far resulted in plausible complication rates, for clinically relevant dose distributions, if $\mathrm{FSU}_{\min} = 0.6$ was chosen. Since clearly not all FSUs are damaged at clinical doses [178], this relatively low $\mathrm{FSU}_{\min}$ seems reasonable.

6.4.5 Inter-patient variation in health status

The response of the rectum of different individuals varies greatly, even when irradiated under the same conditions [184, 185, p. 185]. The first explanation to consider (apart from uncertainties in delivered doses, e.g. due to inter- and intra-fraction movements of the rectum) would be a large inter-patient variation in endothelial cell radiosensitivity. However, there not only seems to be a patient-specific factor (as opposed to treatment-specific) in the development of rectal bleeding for a given *dose level* [68], but also for a given level of *local tissue damage* in terms of telangiectasia [161, 178]. This indicates that the inter-patient difference in overall response of the rectum is not fully explained by a difference in radiosensitivity of the target cell of the tissue; for a given level of local tissue damage some individuals are more prone to develop rectal bleeding than others.

This is not surprising considering that atrophic mucosa, caused by loss of microvasculature, can develop ulcers if exposed to a certain level of faecal mechanical stress, which might be influenced by e.g. the diet [186, 176]. Therefore the tolerated level of local tissue damage varies from patient to patient. This inter-patient variation in the propensity to develop rectal bleeding could be modelled by sampling the value of $\mathrm{FSU}_{\min}$, but like the health status variation for lung this is instead modelled by sampling the pre-treatment FSU density. A value of $\sigma_{FSU} = 10\%$ seems plausible.

6.4.6 Inter-patient variation in radiosensitivity

Above, σ_{FSU} did not represent a variation in health status as much as a general variation in global organ response. Similarly, σ_α will be used to model all variations likely to impact the local tissue response. One of these factors seems to be chance, since the level

of telangiectasia following radiotherapy is related not only to the local dose. Safwat *et al.* [187] explored the stochastic versus deterministic contribution to such inter-patient variation in tissue response by comparing the grade of telangiectasia in two different areas of the skin of each patient, after breast radiotherapy. They found that 10-19% of this variation was due to stochastic factors, and the rest to factors which in principle could be known.

This indicates that, in the rectum in contrast to the lung, the binomial standard deviation in FSU survival is of clinical significance. The magnitude of the binomial standard deviation depends on the number of pre-treatment FSUs per voxel, N_0^{FSU} and the probability of FSU survival, P_{FSU} (given by equation 6.1). At the average dose of Safwat's study, $P_{FSU} \approx 0.77$ with the values of N_{sc}, $\bar{\alpha}$, α/β and N_0^{FSU} chosen above. This gives a standard deviation of 2.9%, which should correspond to 10-19% of the total inter-patient variation in local tissue response, according to reference [187]. Thus, $\sigma_\alpha = 15\%$ will be used, since this corresponds to 84% of the total inter-patient variability in local tissue response, and the remaining 16% is given by the stochastic factor as described above.

Safwat *et al.* supported the hypothesis that the remaining, deterministic, portion of the variation is due to the intrinsic radiosensitivity. Further support for this hypothesis was recently given by the ground-breaking study by Valdagni *et al.* [188] where the risk of rectal bleeding largely depended on the individual's gene profile, in addition to dose-volume parameters for the rectum.

Apart from the variation in intrinsic radiosensitivity and the stochastic factor in FSU inactivation, σ_α might also be influenced by other patient-specific factors such as age and comorbidities [187]. As discussed above, these might be related to the function, and level, of endothelial progenitor cells in the circulation.

6.5 Discussion and Conclusions

6.5.1 Results from the simulations

It was found that the local dose-response of the pulmonary epithelium is relatively steep, with increasing FSU inactivation over a range of 6-22 Gy. The dose-response of the microvascular endothelium in the rectal wall, on the other hand, is much shallower and

with a dose threshold around 35 Gy. These findings, together with considerations of the histology of these tissues, support the hypotheses that the alveolus, and the capillary cluster around a mucosal gland, are the FSUs of the lung (radiation pneumonitis) and rectum (late rectal bleeding) respectively.

In accordance with the consensus in the literature the lung was found to have a large functional reserve, and it seems that the inflammatory process caused by radiation-induced tissue damage holds the key to the volume effect; above a threshold of radiation-induced tissue damage within the CFV, secondary FSU inactivation, likely due to inflammation, develops throughout the CFV and results in organ failure.

The relatively large value of the rectal CFV found in this study is at odds with the common perception of late rectal bleeding as a serial endpoint. However, there are contradictory results concerning the dose range most important for causing rectal bleeding, and the shallow local dose-response curve for local tissue damage indicates that, although large volumes of intermediate doses contribute to the complication probability, high doses are indeed the most important. This could explain the appearance of a small volume effect in some studies, since even a rather small volume of intensive FSU inactivation caused by high doses can result in the number of FSUs in the larger CFV falling below FSU_{min}, but for a different set of dose distributions significantly lower doses to a larger portion of the CFV would have the same effect. In the former case it would be concluded that the volume effect is small and the dose important for bleeding high, whilst in the latter the volume effect would appear larger and the important dose lower.

The parameter values proposed in this chapter are based on the best data available, but as discussed in the next section the radiobiological knowledge base for mechanistic NTCP modelling is not yet complete. In chapter 7 these parameter values were used for generating pseudo-clinical normal-tissue data for lung and rectum, and DVH-based NTCP models were fitted to the resulting datasets. The LKB model parameter values (tables 7.1 and 7.2) differ from recently published meta-analysis data [130, 157]. Although this might partly be due to differences in dose distributions included in the studies, the representation of lung and rectum by the 3D model clearly has limitations.

6.5.2 The radiobiological knowledgebase

Apart from finding parameter values for the 3D model which would result in clinically relevant simulations, the aim of this chapter was to highlight areas which need further research for predictive mechanistic NTCP modelling to be feasible. There is an extensive literature about the effect of radiation on normal tissues, and in spite of the difficulties of producing systematic and reproducible data for normal tissues, great progress has been made; well-founded clinical experience prevents large numbers of patients experiencing serious side-effects from radiotherapy. However, before predictive mechanistic modelling can be developed and used for individualised treatment planning, more specific quantitative data are necessary [74, 33, 21].

It was found that only weak evidence is available for the value of normal-tissue stem cell radiosensitivity parameters, as well as for the identity of the FSU. Also, it is to be expected that the local dose-response of the tissue is influenced not only by these parameters, but also by factors such as comorbidities and the supply of stem cells from the systemic circulation. However, for the purpose of predictive mechanistic modelling, the first step would be to identify the dose-response curve for local tissue damage, with small confidence intervals. If this were verified to be consistent under all relevant conditions, the parameters fitted to this curve, $\bar{\alpha}$ and N_{sc}, could be considered operational without much loss of credibility for the mechanistic model, although future data from radiobiology studies on normal-tissue stem cells might be able to elucidate effects also on this level.

This indicates that, above all, more quantitative histopathological studies on relevant local dose effects are wanted, for great ranges of dose and irradiated volume. It is also of great importance to determine the link between clinical symptoms (e.g. bleeding) and histopathologic conditions (e.g. microvascular malformations). This can be done by a combination of animal models and large clinical studies, where follow-up includes evaluation of the local function/appearance of the tissue in regions which received different dose levels, as well as symptom-oriented questionnaires. It is important that such studies are designed to test specific hypotheses of pathogenesis [157].

Data from such studies could also be used to explore how the build-up of tissue damage over increasing volumes, of different shape, influence the risk of clinical symptoms. Such volume effects could be studied systematically in animal models, and the hypotheses

these generate should be tested in large clinical studies. The success of these studies would depend on accurate knowledge of the dose delivered to small volumes of organs at risk. Therefore current developments in immobilisation, imaging and in vivo dosimetry are encouraging.

Finally, once more data of this kind are available further modelling studies are necessary to put the many pieces of this jigsaw puzzle together and interpret the findings. In this study only one endpoint was considered for each organ, but for some organs, like the rectum, there are several endpoints influencing the quality of life of long-term survivors, which should all be studied similarly.

Chapter 7

The Performance of NTCP Models in the Presence of Confounding Factors

7.1 Introduction

NTCP models have long been used in the analysis of clinical data, and there is a multitude of studies where model parameter values have been fitted to local datasets. In most of these studies a single DVH-based NTCP model is used, but there are a few where a number of models have been fitted to the dataset. In these studies it is generally concluded that the performance of all models is similar [189, 101, 85, 43, 44, 75, 49, 76]. Whether this is due to limitations in the datasets or because no model is better than any of the others is not clear.

One limitation of most datasets, for the purpose of discerning differences between model fits, is the presence of confounding factors, i.e. non-dosimetric factors influencing the outcome independently of the dosimetric variables considered by the NTCP models. Inevitably all models perform less well when applied to datasets with high levels of confounding factors, but it is also expected that any differences in performance between models might be obscured when they cannot be fitted to outcome data relating to the dose distributions only. The importance of considering non-dosimetric factors for predicting the risk of a complication has been demonstrated in several studies [148, 100, 190, 191, 192].

7.1.1 DVH-based modelling

The job of an empirical NTCP model is to "summarise" the information of a dose distribution into a single metric, as efficiently as possible, in order to predict the risk of a complication occuring for a given treatment plan. Ideally the model would predict either 0% or 100% NTCP with a very small confidence interval for all plans, because then a safe plan could be distinguished with certainty from one which would cause a complication. However, this is generally not possible since (1) an empirical NTCP model cannot perfectly summarise the relevant features of a dose distribution, (2) not only the dose distribution but also patient specific factors (confounding factors) influence the outcome. Some of these confounding factors might be known and could in principle be accounted for by the model, but generally they simply influence the model fit to make the slope of the function between the summary measure and NTCP shallower (larger m; see chapter 2).

The parameter characterising the slope of this curve represents the variability in response in the population and is closely connected with the performance of the model. If the curve approaches a step function ($m \to 0$) the chosen dose-effect measure successfully summarises the data and all relevant patient-specific factors are taken into account. In this case the model separates responders from non-responders well, which results in a large area under the ROC curve (AUC) (see chapter 3). Thus it is desirable to find an NTCP model resulting in a high AUC, i.e. a model whose local dose-effect measure summarises the dose distributions so well that the (steep) slope of the response function is due to confounding factors only.

In this chapter DVH-based NTCP models are fitted to pseudo-clinical data generated by the 3D model, for different levels of confounding factors, in order to compare the performance of the models. The parameter values derived for lung and rectum in chapter 6 are used, and to simulate varying levels of confounding factors the values of σ_α and σ_{FSU} are varied. Thus the performance of the fitted models can be explored as a function of the strength of these confounding factors, and the ability of the various models to summarise the dose distributions.

7.1.2 NTCP modelling beyond the dose distribution

Since the performance of NTCP models, and thus their usefulness in clinical practice, is reduced by confounding factors such as inter-patient variations in health status and radiosensitivity, it would be desirable to find methods for removing the influence of these factors. This is of increasing importance the better the models can account for the influence of the dose distribution on the complication probability. In particular, where treatments are planned based on iso-NTCP the dose distribution might not predict the outcome [42], since then all plans are supposedly close to the limit where they would cause a complication, and the patients actually responding with a complication are likely to be amongst the most radiosensitive ones or because of other patient-specific factors be especially susceptible.

In order to move beyond the dose distribution, it is important to identify any non-dosimetric factors influencing the risk of toxicity. However, these factors cannot be accounted for unless methods are developed to classify individuals into risk groups. For some factors, such as smoking status or comorbidities, this should be feasible, whilst for others, such as intrinsic radiosensitivity, there is currently no practical method to find out where on the scale any given patient belongs. As will be shown in this chapter, the potential gain from including relevant non-dosimetric factors into the models motivates research in this area.

If individuals could be stratified according to some risk factor, DVH-based NTCP models could be fitted to subgroups of the larger dataset to obtain parameter values tailored to each patient group. When applied in the clinic the appropriate parameter set would be chosen based on which risk group the patient belonged to. If several risk factors were known the subgroups would be further stratified for each factor. This method would result in steeper response curves for the summary measures, and thus in principle an increase in the AUC measuring the model performance, but it makes inefficient use of the statistics of the dataset, since the models are fitted to small subsets without any input from the other patient groups.

A better method is to account for the non-dosimetric factors in the model itself. This has been done either by including both dosimetric and non-dosimetric variables in a logistic regression model [100, 190], or by the use of a 'dose modifying factor' [193, 191]. In this study, both the logistic regression (logit) function and the cumulative normal

(probit) function will be used with three independent variables influencing the outcome: the summary measure, the radiosensitivity (α) and the health status (N_0^{FSU}). Each of these three factors is known to be important for the outcome in the present datasets, and thus overfitting should not be an issue.

7.2 Methods

7.2.1 Simulations

Pseudo-clinical data were generated with the 3D model of normal-tissue response to radiotherapy, described in chapter 4. In order to represent lung and rectum complication data, the parameter values selected in chapter 6 were used. Each patient was characterised by a unique dose distribution and individually sampled values of α and N_0^{FSU}, representing their personal radiosensitivity and health status. All datasets had different combinations of σ_α and σ_{FSU} but were based on the same set of 1000 dose distributions (separate sets for lung and rectum).

The lung dose distributions were generated as described in sections 4.3.4 and 5.1. In the case of the rectum, beams were directed at the target (prostate) in a four-field box technique, such that the anterior and posterior beams always covered the rectum whilst the portion covered by the fields from the left and right varied from patient to patient. The DVH of the rectum was generated based on the voxels of the rectal wall (excluding the rectal filling). The variation in beam widths and doses was adjusted to make the DVHs resemble rectal DVHs from prostate treatments at Clattrebridge Centre for Oncology, using the methods described in section 5.1.

Both for lung and rectum five datasets were generated: one without confounding factors ($\sigma_\alpha = \sigma_{FSU} = 0$), one with the values of σ_α and σ_{FSU} derived in chapter 6, and three with either σ_α or σ_{FSU} set to zero. Out of these one was generated with σ_α or σ_{FSU} set to the value derived for the other in order to compare the relative effect of these confounding factors on the model performance.

7.2.2 Fitting NTCP models to simulated data

Seven DVH-based NTCP models were fitted to each dataset: the LKB model, the mean dose model, the maximum dose model ($\phi = D_{max}$), the relative seriality model, the

4-parameter critical volume model, the 3-parameter critical volume model and the V_{Dth} model. These models were described in section 2.1. Generally the models were fitted using the maximum likelihood method but when the solutions were unstable, simulated annealing and/or visual inspection of the parameter likelihood space were used to find the best fit.

The performance of the model fits was evaluated in terms of the area under the ROC curve (AUC) when applying the fitted model to the fitting dataset. This assesses the ability of the model to link the outcome to the dose distributions. However, if this were the only criterion for a good model, the choice would be one with many parameters, which would rely on random features in the dataset and thus be much less successful for a different dataset. Therefore, two different model fits were applied to a 'test dataset': the first fitted to a 'fitting dataset', and the other to the test dataset itself. Any case of overfitting would then be detected by finding the AUC for the model fitted to the fitting dataset to be significantly lower than that for the model fitted to the test dataset itself. The fitting and test datasets were different but equivalently generated, and whilst the test dataset consisted of 500 patients, the size of the fitting dataset was gradually reduced from 500 until the difference between the AUC for the two models, applied to the test dataset, was significant.

7.2.3 Models including patient-specific factors

Since the argument of the logistic regression model is the sum of any number of covariates, which can be numerical or categorical, non-dosimetric variables can easily be included. The covariates can be summary measures as well as biological or clinical parameters. A logistic regression-based model [45, 194] calculates the logarithm of the odds of a complication occuring as a sum of the contribution from the n covariates X_i:

$$\ln\left(\frac{NTCP}{1-NTCP}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \tag{7.1}$$

Here β_0 is called the intercept and the magnitude of the parameter β_i depends on the strength of the influence of X_i on the odds. X_1 would typically be the summary measure (ϕ), and $X_2 ... X_n$ any patient-specific variables included in the model. The NTCP is

then given by

$$NTCP = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n)}}. \tag{7.2}$$

The parameters of this model (β_0, β_1, ..., β_n, and any parameters of summary measures) can be fitted to a dataset with the maximum likelihood method (equation 3.2).

Also the cumulative normal function can be used with a set of covariates [194]. With the dose modifying approach mentioned above, non-dosimetric factors are included in the LKB model by adjusting the effective TD_{50} for different subgroups. Thus the EUD response curve is shifted along the EUD axis, whilst using fixed values for the other parameters. However, there is also a method of combining the cumulative normal function with non-dosimetric variables which treats all influencing factors as independent covariates, similarly to the logistic regression model. Here t in equation 2.2 is modified rather than TD_{50}:

$$t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n. \tag{7.3}$$

Alternatively this can be expressed in terms of the usual probit model parameters ϕ_{50} and m:

$$t = \frac{\phi - \phi_{50}}{m \cdot \phi_{50}} + \beta_2 X_2 + ... + \beta_n X_n, \tag{7.4}$$

where β_2, ..., β_n are the model parameters for the non-dosimetric factors $X_2, ..., X_n$. These two expressions for t are related with the following parameter transformation:

$$\begin{cases} \phi_{50} = -\beta_0/\beta_1 \\ m = -1/\beta_0 \end{cases}.$$

7.3 Results

7.3.1 The influence of confounding factors

In the overfitting test the difference in AUC for the models fitted to the fitting dataset and the test dataset respectively was statistically significant in only one case: the LKB model fitted to the first lung dataset. However, this difference in AUC was borderline and only appeared in the third AUC decimal. Thus the AUC for the fitted models applied to the fitting dataset itself could be used as a measure of model performance.

Figures 7.1 and 7.2 show the performance of the DVH-based NTCP models when

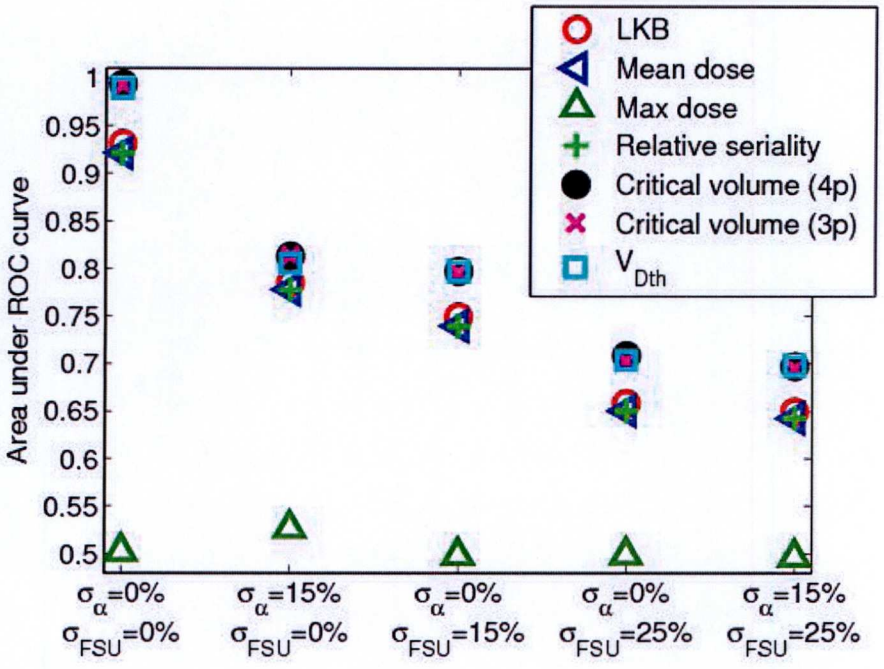


Figure 7.1: The performance (AUC) of seven DVH-based models fitted to pseudo-clinical lung datasets which were simulated with different levels of confounding factors, in terms of the variation in radiosensitivity σ_α and health status σ_{FSU}.

fitted to the lung and rectum datasets respectively, with increasing levels of confounding factors. Since the models are based only on the DVHs they cannot account for individual values of radiosensitivity or health status, which explains the decreasing model performance with increasing confounding factors.

The AUC was found to be influenced by the NTCP values (and thus the complication rate) in the dataset which the model fits were applied to; it was more difficult for the models to correctly classify DVHs with NTCP $\sim 50\%$ than DVHs with low or high NTCP, and thus performed worse on a dataset with many DVHs with NTCP $\sim 50\%$. This effect might influence the results in figures 7.1 and 7.2 since the datasets have increasing complication rate with increasing levels of confounding factors. However, when similar datasets were generated with parameter values adjusted to produce similar complication rates, the results still showed a decrease in AUC with increasing levels of confounding factors.

All DVH-based models fitted to the datasets, except the maximum dose model, performed similarly well, and the influence of the confounding factors on the model performance was more important than the choice of model. The superiority of the mean dose model compared to the maximum dose model, for both lung and rectum data, is

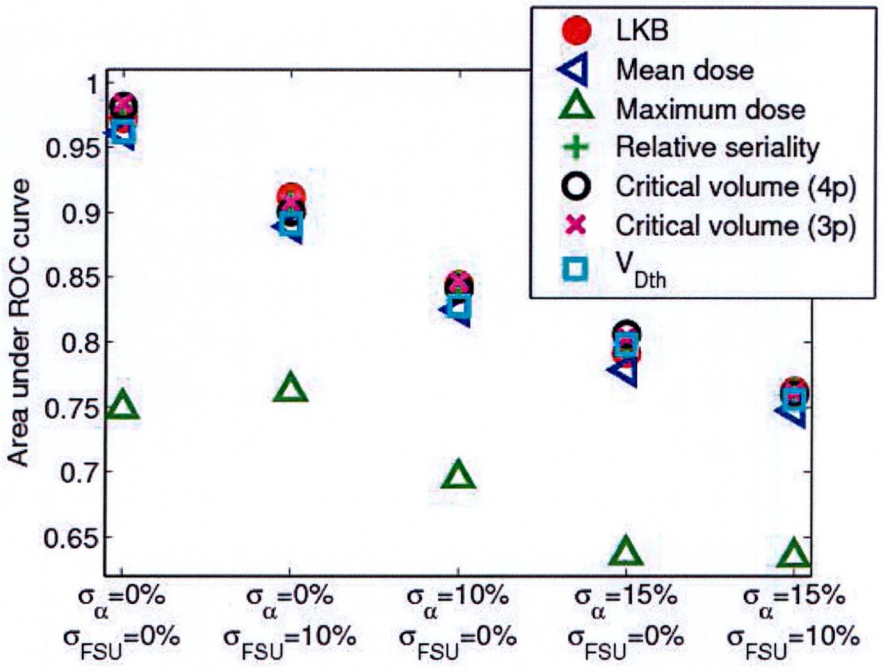


Figure 7.2: The performance (AUC) of seven DVH-based models fitted to pseudo-clinical rectum datasets which were simulated with different levels of confounding factors, in terms of the variation in radiosensitivity σ_α and health status σ_{FSU}.

consistent with the high values of the LKB volume parameter n, and the low values of the relative seriality parameter s, in tables 7.1 and 7.2 (at the end of the chapter) which show all parameter values for the model fits shown in figures 7.1 and 7.2.

In the first column of tables 7.1 and 7.2 the cumulative normal function parameter, ϕ_{50}, represents TD_{50} for the LKB model, the mean dose resulting in 50% complication rate for the mean dose model, the maximum dose resulting in 50% complication rate for the maximum dose model, and rdV_{50} for the 4-parameter critical volume, 3-parameter critical volume and V_{Dth} models. In general, ϕ_{50} is the value of the model-specific summary measure of the DVH, at NTCP = 50%. For the relative seriality model, which is not based on the cumulative normal function, D_{50} is listed in the first column, γ in the second column and s in the third column. For the other models the cumulative normal function 'slope' parameter, m, is listed in the second column. For the 3- and 4-parameter models, model-specific parameters relating to the DVH reduction into a summary measure are given in columns 3 and 4.

As expected the value of the slope parameter reflected the amount of inter-patient variation in the datasets; in both tables m increased and γ decreased with increasing variation. The values of ϕ_{50} and D_{50} were generally consistent amongst the datasets,

except the unrealistically high ϕ_{50} for the maximum dose model in table 7.1 which confirms that this is an unsuitable model for lung datasets.

For the lung datasets (table 7.1), ϕ_{50} of the LKB model was lower than ϕ_{50} for the mean dose model, and n was consistently greater than one. This means that the LKB model found lower doses to be important (or the volume receiving lower doses), which is in line with the suggestion in the next section that critical volume type models might be more suitable for lung datasets. The performance of the relative seriality model was very similar to the mean dose model, and table 7.1 shows that ϕ_{50}/D_{50} was very similar for these models.

For the rectum datasets (table 7.2), n had intermediate values. Thus the LKB model fits correctly identified a smaller volume effect for the simulated rectum than for lung. In contrast the relative serialiy model fits resulted in surprisingly low values of s, indicating a large volume effect, and D_{50} was lower than ϕ_{50} of the LKB model, but higher than ϕ_{50} of the mean dose model. Interestingly, ϕ_{50} and m of the critical volume type models were similar to the corresponding parameters fitted to the lung datasets, though D_{50} was higher.

Whilst D_{50} for lung had a value causing very high FSU inactivation (cf. figure 6.3), D_{50} for rectum corresponded to local doses causing very low levels of FSU inactivation (cf. figure 6.9), except for the V_{Dth} model fitted to the last rectum dataset. In a few cases the probaility distribution of the parameters showed two peaks when fitted to rectum datasets, with a low D_{50} and high ϕ_{50} or a high D_{50} and a low ϕ_{50} respectively. However, in most cases the probability of $D_{50} \sim 35$ was higher than $D_{50} \sim 63$, as shown by the maximum likelihood estimates in table 7.2. This shows that although the parameters of the critical volume type models have a biological interpretation, as empirical models they cannot be assumed to represent the true dose volume relationship.

7.3.2 The relative model performance

Although the performance of most models was similar, for the lung datasets the three critical volume type models (the V_{Dth} and the 3- and 4-parameter critical volume models) were consistently superior ($p < 0.05$, except V_{Dth} compared to LKB for the second dataset with $p = 0.08$). For the simulated rectum datasets, on the other hand, both types of models performed equally well. The reason for this could be that the outcome

for the lung simulations is predominantly linked to the damaged volume, due to the large CFV, and thus better predicted by the critical volume models which reduce the DVH to a volume; whilst the intermediate CFV of the rectum could make the spatial distribution of damage especially important, which is lost in all DVH reduction schemes (see related discussion in section 5.7).

The pattern of the difference in AUC between the critical volume type models and the other three well-performing models (the LKB, mean dose and relative seriality models) was studied in more detail, as a function of σ_{FSU}. This difference was expected to decrease as AUC approached 0.50, as the confounding factors grew stronger. For values of σ_{FSU} between 0 and 50% ($\sigma_\alpha = 0$), models were fitted to five simulated 200-patient datasets, in order to provide information about the variation in model performance for different datasets, and thus the uncertainty in the difference between the model groups.

Figure 7.3 (a) shows the actual mean AUC at each σ_{FSU}, for all models in the comparison, and figure 7.3 (b) shows the *difference* in mean AUC between the critical volume type models and the other three well-performing models, together with the standard error of the difference, based on the five datasets. It was found that rather than continuously decreasing, the difference in model performance was relatively large (0.07) below a threshold of the level of confounding factors, and smaller (0.03) above the threshold. Since the detection of a difference between types of models depends on there being enough statistical power in the dataset to select the ideal summary measure (which is made difficult by cross-correlation between dose-volume parameters), these results indicate that this statistical power is lost rather suddenly, compared to the continually decreasing correlation between model prediction and outcome with increasing σ_{FSU}. Thus, although models have performed equally well for real clinical datasets so far, differences might be discovered if confounding factors are taken into account.

7.3.3 Model performance after stratifying for patient-specific factors

Clinical outcome data inevitably include significant levels of confounding factors. However, if the radiosensitivity and health status (and any other important non-dosimetric factors) were known for all patients, and could be quantified, the dataset could be stratified for such patient-specific factors, and the models fitted separately to each

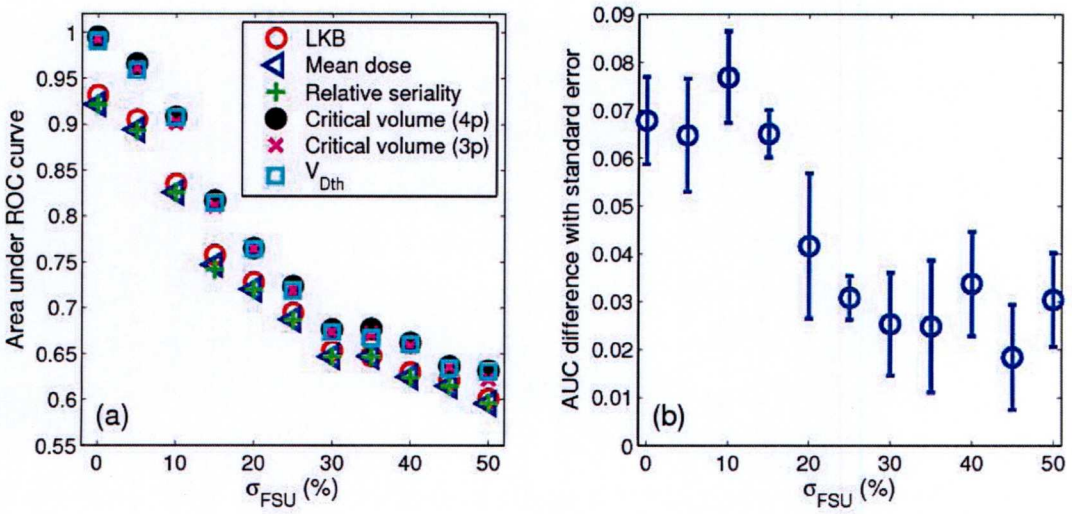


Figure 7.3: (a) Mean AUC for six well-performing models for 5 equivalently generated datasets of 200 patients each, at different levels of σ_{FSU}. (b) The difference in average model performance (mean AUC), for simulated lung data, between two groups of DVH-based NTCP models: three critical volume type models (the V_{Dth} and the 3- and 4-parameter critical volume models) and the other three models (the LKB, mean dose and relative seriality models).

subset. This would lead to a lower level of confounding factors since in each subset all patients would have similar radiosensitivity and health status. DVH-based models would then be expected to perform better for each subset compared to the total dataset. However, the parameter values would be different for each subset, and the right parameter set would have to be chosen when applying the model to a new patient.

This hypothesis was tested by stratifying dataset 5, which included inter-patient variations in both radiosensitivity and health status, into four groups of similar α, and these groups into four further subgroups of similar N_0^{FSU}. The models were then fitted to each of the 16 subgroups and their performance evaluated. The results for two of the models, fitted to the simulated lung dataset 5, are presented in table 7.3 (LKB model) and table 7.4 (3-parameter critical volume model). Similarly, tables 7.5 and 7.6 show the AUCs for these models fitted to the simulated rectum dataset 5.

As expected, a great improvment was seen in some of the subgroups, compared to when the models were fitted to the total dataset. For example, when the LKB model was fitted to the total dataset of table 7.3 (dataset 5 in figure 7.1) the AUC was 0.65, whilst for the corresponding dataset without confounding factors (dataset 1 in figure 7.1) it was 0.93. Most of the values in table 7.3 are within this range, and some close to

the upper limit. A certain variability in the patient-specific factors still exists within the subsets, since each variable was divided into only four bins. This is reflected by the values of the AUC generally being lower than that for the dataset without confounding factors.

As mentioned in section 7.1.2, a great disadvantage of splitting the dataset into subsets before fitting the models is the loss of statistical power. Many of the subgroups didn't have enough responders to give reliable model parameters, and where no or all patients suffered the complication the models could not be fitted at all. Actually including the patient-specific factors in the models is much more efficient since, instead of fitting three parameters to e.g. sixteen different subsets (48 paramter values in total), five parameters (three dosimetric and two non-dosimetric) can be fitted to the total dataset, and all information from the data is considered when fitting each parameter. This method was explored below.

7.3.4 Including non-dosimetric variables

So far only the DVHs were considered by the models fitted to the simulated data. For most models this was done by reducing the DVH to a summary measure and linking it to NTCP via the probit function. In order to also account for the patients' radiosensitivity and health status, α and N_0^{FSU} were now introduced as covariates alongside the model-specific ϕ (see section 7.2.3). This could not be done for the relative seriality model, which has a different structure.

As in the previous section, dataset 5 for both the simulated lung and rectum data were used, which included clinically relevant levels of the confounding factors. The different models were adapted to include patient-specific factors using both the logit function and the probit function, as described in section 7.2.3. Tables 7.7 and 7.8 show the maximum likelihood fitted parameter values. β_0 is the intercept, and β_1, β_2 and β_3 the coefficients associated with ϕ, the radiosensitivity and the health status respectively. The other parameters are the model specific parameters required to calculate ϕ. In order to facilitate optmisation the health status was represented by $N_0^{FSU}/10,000$, giving values more similar to the radiosensitivity parameter α. The positive values of β_1 and β_2 indicate that higher values of ϕ and α increase the risk of complications, whilst the negative value of β_3 indicates an increase in the risk of complications for decreasing N_0^{FSU}.

Figures 7.4 (a) and (b) show the performance of these models and, for comparison, the performance of the corresponding standard models for the datasets without confounding factors (first datapoints in figures 7.1 and 7.2).

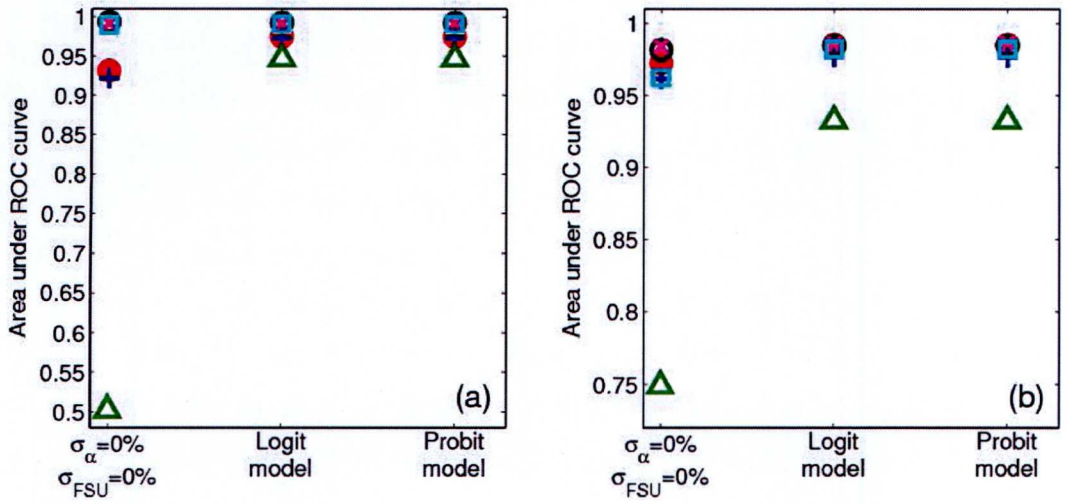


Figure 7.4: Performance of the logit and probit models including patient-specific factors for (a) lung dataset 5 and (b) rectum dataset 5, compared to the performance of the models for the dataset without confounding factors. The models, which included different model-based summary measures, were the LKB model (•), the mean dose model (+), the maximum dose model (△), the 4-parameter critical volume model (○), the 3-parameter critical volume model (×) and the V_{Dth} model (□).

As expected the performance improved greatly compared to when only the DVH was considered for the same datasets (dataset 5 in figures 7.1 and 7.2), reaching levels similar to the performance for datasets without confounding factors. Interestingly, the poorest models performed better on dataset 5 when patient-specific factors were considered, than for dataset1 which did not include confounding factors. This does not indicate that the maximum dose model was a very good summary measure for these datasets, but rather that including more predictors made it possible for the model to classify more patients correctly, since now not only the dose distribution influenced the outcome.

The performance of the models based on the logit function and the probit function were essentially identical.

7.4 Discussion

7.4.1 The AUC as a measure of model performance

In this chapter, the performance of the DVH-based models has been compared in terms of the AUC for the fitting dataset. The AUC is the probability of a responder having a higher NTCP than a non-responder when one of each kind is randomly picked from the dataset, and thus measures how well the model rank-orders the DVHs. As mentioned in section 7.3.1, the value of the AUC depends on the magnitude of the NTCPs in the dataset, since there are more errors in the rank-ordering of a dataset where many DVHs have NTCP $\sim 50\%$, than where most plans have an NTCP of 0 or 100%.

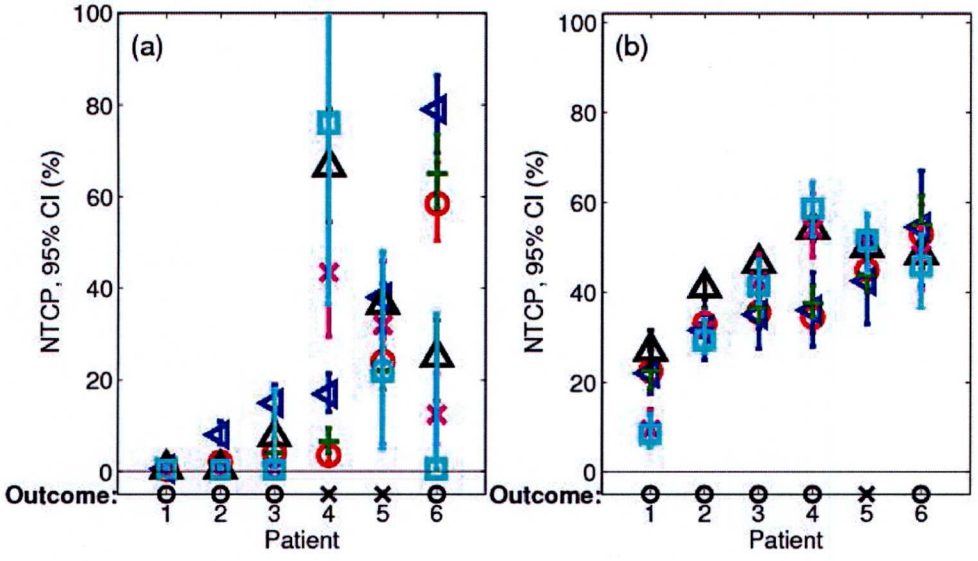


Figure 7.5: NTCP estimates with 95% confidence intervals (CI) for six patients from (a) lung datasets 1 and (b) lung dataset 5. Each patient has the same dose distribution in both figures, but whilst there are no confounding factors in the dataset in (a), values for the radiosensitivity and health status were randomly sampled when simulating the dataset in (b). The outcome is given on the abscissa as complication ($\times$) or no complication ($\circ$). The figures show the NTCP estimates by the LKB model ($\circ$), the mean dose model ($\triangleleft$), the relative seriality model ($+$), the 4-parameter critical volume model ($\triangle$), the 3-parameter critical volume model ($\times$) and the V_{Dth} model ($\square$).

This is illustrated by figures 7.5 (a) and (b) which present the predicted NTCP for a selection of patients in two datasets, the first without confounding factors (lung dataset 1), and the second with clinically plausible levels of confounding factors (lung dataset 5). Here the high AUC of the models when fitted to the first dataset is reflected by the few intermediate NTCP values in (a), compared to (b), for which dataset the models had a much lower AUC.

Similarly, a model which correctly assigns DVHs predominantly very low or very high NTCPs, reaches a higher AUC than a model with many intermediate NTCP values for the same dataset. This could be seen, for example in lung dataset 1, where more patients had intermediate NTCP estimates with the LKB model than with the 3-parameter critical volume model, which had a higher AUC; see figures 7.6 (a) and (b).

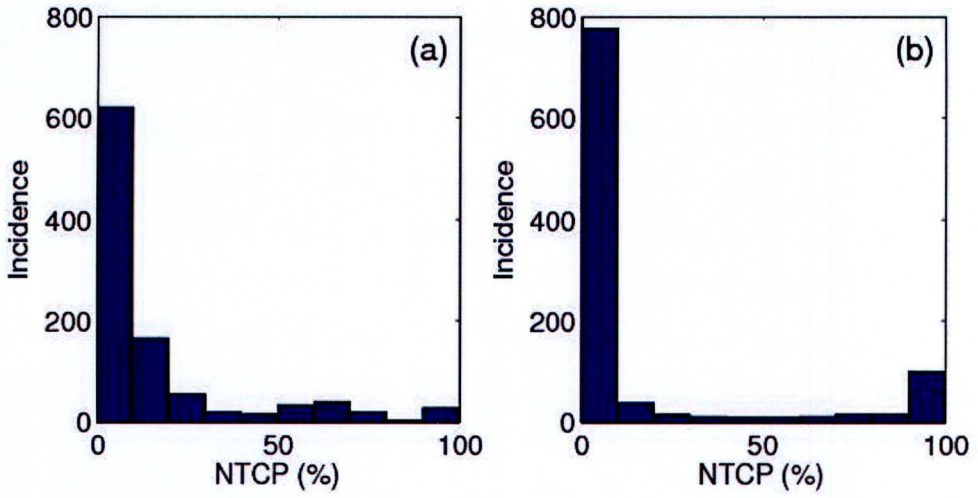


Figure 7.6: Histograms of predicted NTCPs for lung dataset 1 by (a) the LKB model and (b) the 3-parameter critical volume model. The lower AUC for the LKB model is reflected by more patients having intermediate NTCP values.

It can be concluded that, since a model which predominantly assigns very low or very high NTCP values also labels plans as low- or high-risk plans with great certainty, the AUC is a suitable measure of model performance for clinical data. A 'good' model correctly accounts for all relevant factors influencing the outcome; on the other hand, even a suboptimal model can be considered validated if the AUC is as high for testing datasets as for the fitting dataset. The following graphical methods are sometimes used to demonstrate the validity of a model.

7.4.2 Graphical methods of evaluating model performance

The performance of NTCP models can be evaluated graphically by binning the data into groups of similar ϕ and comparing the complication rate in the bins with the NTCP as a function of ϕ. Figures 7.7 and 7.8 show such plots for dataset 1 and 5 for lung, which resulted in an average AUC of 0.96 and 0.67 respectively. The lower AUCs for dataset 5, which included high levels of confounding factors, are reflected by the much shallower slope of the predicted NTCP curves in figure 7.8. Note that the ranges of the ϕ-axes

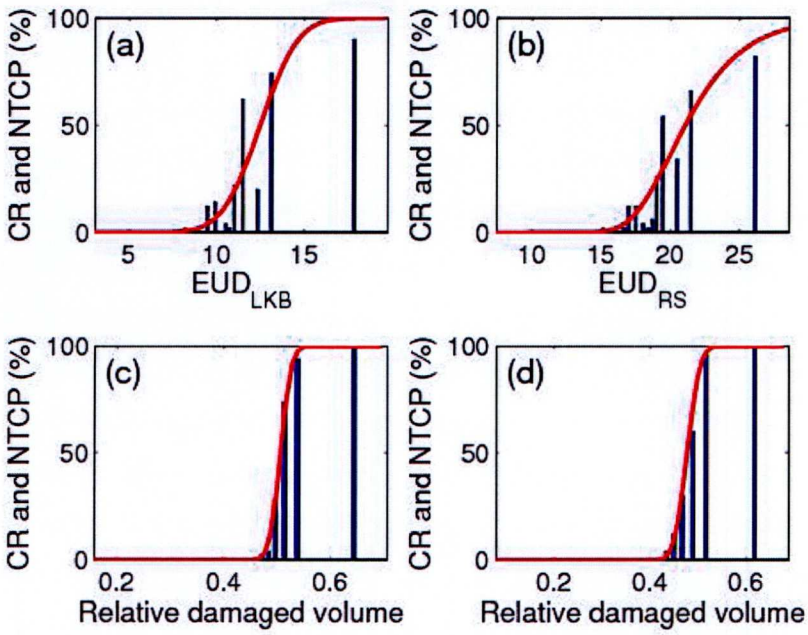


Figure 7.7: Fitted NTCP models compared to the complication rate (CR) in bins of 50 patients with similar EUD or relative damaged volume, for the first lung dataset ($\sigma_\alpha = 0\%$ and $\sigma_{FSU} = 0\%$). The fitted models are (a) the LKB model, (b) the relative seriality model, (c) the 3-parameter critical volume model, and (d) the V_{Dth} model.

were normalised to the range of ϕ-values in the dataset, with constant proportional margins from the lowest and highest bin values to the frames. This makes it possible to compare the steepness of the curves, in spite of the different units of the ϕ axes. There are twenty bins in each figure, but in some cases the majority are invisible, due to a complication rate of zero in these bins.

For the four 3-parameter models (LKB, relative seriality, critical volume and V_{Dth}) shown in the figures, NTCP is a function of the equivalent uniform dose ($\mathrm{EUD_{LKB}}$ or $\mathrm{EUD_{RS}}$), or the relative damaged volume. The expression for $\mathrm{EUD_{RS}}$, based on equation 2.12, depends on two of the fitted parameters and the predicted NTCP:

$$\mathrm{EUD_{RS}} = D_{50}\left(1 - \frac{\ln(\ln(1/\mathrm{NTCP})) - \ln(\ln 2)}{e\gamma}\right) \tag{7.5}$$

Figure 7.7 illustrates the difference in model performance when the AUC were 0.93 ((a) and (b)) and 0.99 ((c) and (d)). As well as having steeper curves the critical volume models clearly manage to summarise the dose distributions better, since the complication rate increases monotonically with the relative damaged volume. This effect is less apparent in the presence of confounding factors (figure 7.8).

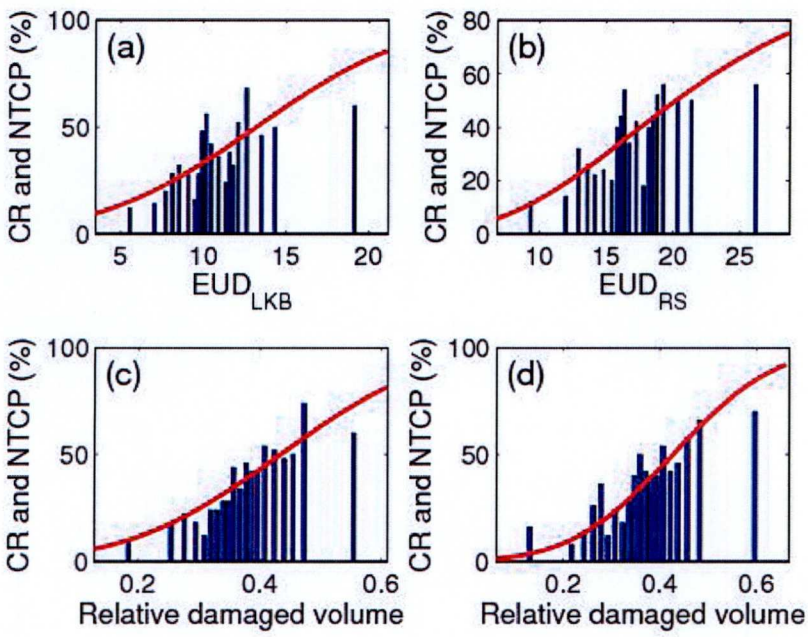


Figure 7.8: Fitted NTCP models compared to the complication rate (CR) in bins of 50 patients with similar EUD or relative damaged volume, for the fifth lung dataset ($\sigma_\alpha = 15\%$ and $\sigma_{FSU} = 25\%$). The fitted models are (a) the LKB model, (b) the relative seriality model, (c) the 3-parameter critical volume model, and (d) the V_{Dth} model.

Figure 7.9 shows another graphical evaluation of the model performance on the two lung datasets used in this section. Here, the datasets are binned according to their predicted NTCP, and the NTCP of the bins is compared to the actual complication rate in the bins. The predictions of a good model should be close to the $y = x$ line.

Both the previous graphical performance evaluation and the present suffer from having to consider the average NTCP estimates in data subgroups, since these are compared to the complication rate which can only be calculated for a group of patients. Therefore, these methods are not as effective as the AUC where the classification of every single plan influences the performance measure. With these graphical methods a model which identified the right fraction of DVHs in each subgroup as high-risk, would get a good performance estimate, even though the wrong individual plans had been identified. The AUC, on the other hand, would detect the failure of the model to classify individual plans.

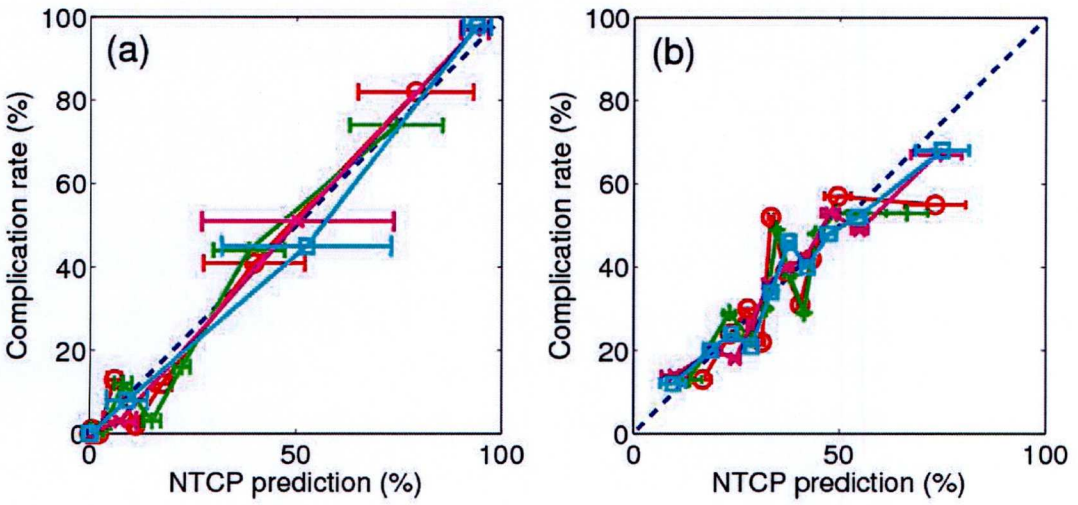


Figure 7.9: Patients binned according to predicted NTCP by the LKB (○), relative seriality (+), critical volume (×) and V_{Dth} (□) models, and plotted against the actual complication rate in the groups (100 patients/group). The dashed line represents $y = x$. The datasets are for lung, (a) without confounding factors and (b) with $\sigma_\alpha = 15\%$ and $\sigma_{FSU} = 25\%$, corresponing to dataset 1 and 5 in figure 7.1.

7.4.3 The relative performance of NTCP models

This study shows that the performance of NTCP models is influenced more by confounding factors than by the choice of NTCP model. This is in agreement with a recent study fitting three models to a large dataset of rectal complications following prostate radiotherapy [195]; all models performed similarly but including clinical factors significantly improved the fit. However, two of the models were based on the EUD: the LKB model and a logistic model with EUD as summary measure. Thus they were essentially the same model and the third model, the relative seriality, was shown to perform very similarly to the LKB model (which is in agreement with the findings in this chapter). Greater differences in model performance might have been found if also a critical volume type model was included in the study.

The results in this chapter do not answer the question of which of the DVH-based models is best. The summary measure of the critical volume model might be expected to be the most successful one, since it explicitly and transparently takes both dose and volume into account, by estimating the damaged volume based on the local dose. However, the performance of an empirical model also depends on how easily it can be fitted to a dataset, and it has been reported that one of the four parameters of the critical volume model cannot be determined with accuracy from clinical datasets [60, 43, 44].

Also, as shown by van Luijk *et al.* [60] strong correlations between model parameters make it difficult to determine confidence intervals for the parameter values.

For the datasets and models considered in this work, best fits could be found, though in some cases the upper confidence limit for a parameter was infinite. On a practical note, the mean dose model and the 4-parameter critical volume models had the advantage of easily converging to the same optimal solution, using the maximum likelihood method, for different starting points in the parameter space. The V_{Dth} model was quite unstable and the starting point often had to be close to the optimal solution for the maximum likelihood method to find it. This was dealt with by using the more time consuming simulated annealing approach, as well as a visual inspection of the full parameter probability distribution. Creating the probability distributions was very computationally expensive though, and this was usually done by running parts of the job in parallel on the Condor pool (see section 4.3) at the University of Liverpool.

Interestingly, the LKB model and the relative seriality models often performed very similarly; see discussion in section 2.1.3. However, fitting the relative seriality model to a dataset took a lot longer than fitting the LKB model.

7.4.4 The influence of confounding factors

In this work only three selected variables which were known to influence the outcome (ϕ, α and N_0^{FSU}) were included, when modelling also non-dosimetric factors. This is in contrast to the approach by the Washington University group, which is based on an automated variable selection which allows models built from several dose-volume parameters as well as patient-specific variables [100]. With their approach the dataset is heavily relied on for inference as well as parameter fitting; no *a priori* understanding of the cause and effect of the complication of interest is taken into account in the variable selection. Thus the resulting models are optimised for the characteristics of the fitting dataset but it is likely that the specific combination of dose-volume variable selected by this method would be different for other datasets.

This leads to the question whether a good model should yield high AUC values for all datasets; depending on the interpatient variablility in dosimetric and non-dosimetric variables, these variables are not of the same relative importance for a predictive model applied to different datasets. The fact that the most important predictive variables for

a given endpoint differ between different studies could be explained by some variables being effectively constant from patient to patient in one dataset, whilst varying in another. Or, as observed by Dehing-Oberije *et al.* [148], dosimetric variables seem to be more important for datasets including treatments of high mean lung dose than for datasets with only low mean lung doses.

However, as shown by the results in section 7.3.4, correctly accounting for the most important predictive variables (nondosimetric in dataset 5) results in a high AUC, irrespective of how well(/poorly) less important variables are accounted for (e.g. maximum dose in dataset 5). The interpatient variability in dosimetric and non-dosimetric variables are likely to differ between datasets from different centres and years of treatment, and a general predictive model could be built by identifying, and quantifying the influence of, such variables from the datasets where a large variability exists. Thus, the general model might include a large number of variables, but only some of them might be relevant for a given dataset. The AUC would be high only if at least all the relevant variables were selected into the model.

Figure 7.1 showed that for lung the relative importance of σ_α and σ_{FSU} was comparable. For the rectum datasets on the other hand, figure 7.2 showed that the performance of all models was influenced more by σ_α than by σ_{FSU}. The reason for this is not clear, but could be related to the smaller volume effect of the rectum compared to the lung; organs with a small functional reserve are especially vulnerable to severe local tissue damage, whether due to excessive dose or high radiosensitivity. However, it might also reflect characteristics in the dataset.

7.5 Conclusions

The results in this chapter indicate that the influence of confounding factors on the performance of NTCP models is greater than the choice of model. The great improvement in the performance of the models when relevant non-dosimetric factors were included, motivates efforts to find methods to identify and quantify such factors.

Once relevant patient-specific factors are known and can be included in the NTCP models, critical volume models could prove to be superior to other DVH-based NTCP models for lung. However, efforts should first be made to account for confounding factors.

The most efficient way to account for non-dosimetric variables is to include them as covariates alongside a (DVH) summary measure in a logit function- or probit function based model.

Table 7.1: Maximum likelihood estimates of the parameter values for fits of NTCP models to pseudo-clinical lung data presented in figure 7.1. See text for definitions of parameters.

	ϕ_{50}/D_{50}	m/γ	$n/s/D_{50}$	k
dataset 1: $\sigma_\alpha = 0\%$, $\sigma_{FSU} = 0\%$				
LKB	12.5 Gy	0.136	3.01	
Mean dose	20.6 Gy	0.112		
Maximum dose	638 Gy	0.869		
Relative seriality	21.1 Gy	2.71	7.78×10^{-4}	
Critical volume (4p)	0.465	0.0305	13.6 Gy	10.3
Critical volume (3p)	0.507	0.0313	15.1 Gy	
V_{Dth}	0.478	0.0429	14.1 Gy	
dataset 2: $\sigma_\alpha = 15\%$, $\sigma_{FSU} = 0\%$				
LKB	15.8 Gy	0.261	1.61	
Mean dose	20.3 Gy	0.230		
Maximum dose	115 Gy	0.663		
Relative seriality	20.1 Gy	1.60	4.10×10^{-4}	
Critical volume (4p)	0.464	0.189	13.3 Gy	6.79
Critical volume (3p)	0.531	0.182	13.9 Gy	
V_{Dth}	0.476	0.243	14.0 Gy	
dataset 3: $\sigma_\alpha = 0\%$, $\sigma_{FSU} = 15\%$				
LKB	13.2 Gy	0.320	2.50	
Mean dose	20.6 Gy	0.271		
Maximum dose	413 Gy	1.36		
Relative seriality	20.7 Gy	1.29	9.42×10^{-5}	
Critical volume (4p)	0.459	0.254	14.4 Gy	31.9
Critical volume (3p)	0.573	0.435	13.8 Gy	
V_{Dth}	0.459	0.256	14.7 Gy	
dataset 4: $\sigma_\alpha = 0\%$, $\sigma_{FSU} = 25\%$				
LKB	12.4 Gy	0.555	2.69	
Mean dose	19.9 Gy	0.471		
Maximum dose	280 Gy	2.31		
Relative seriality	19.9 Gy	0.800	5.15×10^{-5}	
Critical volume (4p)	0.505	0.304	11.5 Gy	58.9
Critical volume (3p)	0.473	0.354	15.6 Gy	
V_{Dth}	0.422	0.470	15.3 Gy	
dataset 5: $\sigma_\alpha = 15\%$, $\sigma_{FSU} = 25\%$				
LKB	13.2 Gy	0.562	2.34	
Mean dose	20.1 Gy	0.481		
Maximum dose	281 Gy	2.22		
Relative seriality	20.2 Gy	0.787	8.04×10^{-5}	
Critical volume (4p)	0.427	0.446	15.0 Gy	46.4
Critical volume (3p)	0.434	0.449	14.9 Gy	
V_{Dth}	0.427	0.387	16.6 Gy	

Table 7.2: Maximum likelihood estimates of the parameter values for fits of NTCP models to pseudo-clinical rectum data presented in figure 7.2. See text for definitions of parameters.

	ϕ_{50}/D_{50}	m/γ	$n/s/D_{50}$	k
dataset 1: $\sigma_\alpha = 0\%$, $\sigma_{FSU} = 0\%$				
LKB	46.4 Gy	0.0433	0.466	
Mean dose	38.7 Gy	0.0651		
Maximum dose	78.7 Gy	0.0683		
Relative seriality	40.2 Gy	8.20	2.96×10^{-5}	
Critical volume (4p)	0.458	0.0770	34.7 Gy	8.84
Critical volume (3p)	0.535	0.0564	35.6 Gy	
V_{Dth}	0.481	0.115	34.6 Gy	
dataset 2: $\sigma_\alpha = 0\%$, $\sigma_{FSU} = 10\%$				
LKB	51.8 Gy	0.0607	0.293	
Mean dose	38.6 Gy	0.109		
Maximum dose	76.4 Gy	0.0606		
Relative seriality	42.4 Gy	4.53	8.51×10^{-3}	
Critical volume (4p)	0.456	0.161	34.8 Gy	8.91
Critical volume (3p)	0.524	0.122	36.2 Gy	
V_{Dth}	0.538	0.224	31.0 Gy	
dataset 3: $\sigma_\alpha = 10\%$, $\sigma_{FSU} = 0\%$				
LKB	50.8 Gy	0.0974	0.335	
Mean dose	39.3 Gy	0.152		
Maximum dose	77.7 Gy	0.0833		
Relative seriality	43.7 Gy	3.36	0.0370	
Critical volume (4p)	0.462	0.221	35.5 Gy	8.16
Critical volume (3p)	0.548	0.172	35.4 Gy	
V_{Dth}	0.548	0.283	31.1 Gy	
dataset 4: $\sigma_\alpha = 15\%$, $\sigma_{FSU} = 0\%$				
LKB	48.7 Gy	0.134	0.391	
Mean dose	39.0 Gy	0.391		
Maximum dose	78.6 Gy	0.113		
Relative seriality	43.2 Gy	2.75	0.0686	
Critical volume (4p)	0.475	0.266	33.3 Gy	9.53
Critical volume (3p)	0.545	0.210	35.0 Gy	
V_{Dth}	0.528	0.323	31.1 Gy	
dataset 5: $\sigma_\alpha = 15\%$, $\sigma_{FSU} = 10\%$				
LKB	50.5 Gy	0.152	0.341	
Mean dose	39.0 Gy	0.220		
Maximum dose	78.0 Gy	0.118		
Relative seriality	45.9 Gy	2.51	0.165	
Critical volume (4p)	0.443	0.336	36.4 Gy	10.1
Critical volume (3p)	0.522	0.264	37.3 Gy	
V_{Dth}	0.340	0.393	62.5 Gy	

Table 7.3: AUC of the LKB model fitted to subsets of lung dataset 5, stratified according to radiosensitivity (α) and health status, represented by the initial FSU density (N_0^{FSU}). The complication rate in each subset is given in brackets. The values of α and N_0^{FSU} are the middle values of the bins. The AUC for the total dataset was 0.65, and for the corresponding dataset without confounding factors 0.93.

	$\alpha = 0.19$	$\alpha = 0.23$	$\alpha = 0.25$	$\alpha = 0.33$
$N_0^{FSU} = 2200$	0.61 (25/62)	0.82 (40/62)	0.83 (59/62)	- (62/62)
$N_0^{FSU} = 3300$	1.0 (1/62)	0.74 (19/62)	0.80 (36/62)	0.94 (54/62)
$N_0^{FSU} = 3900$	- (0/62)	0.61 (3/62)	0.95 (13/62)	0.79 (32/62)
$N_0^{FSU} = 5800$	- (0/62)	- (0/62)	1.0 (1/62)	0.90 (13/62)

Table 7.4: AUC of the 3-parameter critical volume model fitted to subsets of lung dataset 5, stratified according to radiosensitivity (α) and health status, represented by the initial FSU density (N_0^{FSU}). The complication rate in each subset is given in brackets. The values of α and N_0^{FSU} are the middle values of the bins. The AUC for the total dataset was 0.69, and for the corresponding dataset without confounding factors 0.99.

	$\alpha = 0.19$	$\alpha = 0.23$	$\alpha = 0.25$	$\alpha = 0.33$
$N_0^{FSU} = 2200$	0.77 (25/62)	0.84 (40/62)	0.95 (59/62)	- (62/62)
$N_0^{FSU} = 3300$	1.0 (1/62)	0.92 (19/62)	0.97 (36/62)	0.96 (54/62)
$N_0^{FSU} = 3900$	- (0/62)	0.90 (3/62)	0.95 (13/62)	0.79 (32/62)
$N_0^{FSU} = 5800$	- (0/62)	- (0/62)	1.0 (1/62)	0.91 (13/62)

Table 7.5: AUC of the LKB model fitted to subsets of rectum dataset 5, stratified according to radiosensitivity (α) and health status, represented by the initial FSU density (N_0^{FSU}). The complication rate in each subset is given in brackets. The values of α and N_0^{FSU} are the middle values of the bins. The AUC for the total dataset was 0.76, and for the corresponding dataset without confounding factors 0.97.

	$\alpha = 0.038$	$\alpha = 0.048$	$\alpha = 0.052$	$\alpha = 0.070$
$N_0^{FSU} = 310$	- (0/62)	0.93 (9/62)	0.86 (33/62)	0.91 (43/62)
$N_0^{FSU} = 360$	- (0/62)	0.92 (3/62)	0.88 (13/62)	0.97 (36/62)
$N_0^{FSU} = 380$	- (0/62)	- (0/62)	0.91 (7/62)	0.93 (33/62)
$N_0^{FSU} = 450$	- (0/62)	0.98 (2/62)	0.98 (2/62)	0.82 (12/62)

Table 7.6: AUC of the 3-parameter critical volume model fitted to subsets of rectum dataset 5, stratified according to radiosensitivity (α) and health status, represented by the initial FSU density (N_0^{FSU}). The complication rate in each subset is given in brackets. The values of α and N_0^{FSU} are the middle values of the bins. The AUC for the total dataset was 0.76, and for the corresponding dataset without confounding factors 0.98.

	$\alpha = 0.038$	$\alpha = 0.048$	$\alpha = 0.052$	$\alpha = 0.070$
$N_0^{FSU} = 310$	- (0/62)	0.95 (9/62)	0.93 (33/62)	0.90 (43/62)
$N_0^{FSU} = 360$	- (0/62)	1.0 (3/62)	0.94 (13/62)	0.97 (36/62)
$N_0^{FSU} = 380$	- (0/62)	- (0/62)	0.92 (7/62)	0.93 (33/62)
$N_0^{FSU} = 450$	- (0/62)	1.0 (2/62)	1.0 (2/62)	0.81 (12/62)

Table 7.7: Maximum likelihood estimates of the parameter values for fits of NTCP models including confounding factors to pseudo-clinical lung dataset 5.

	β_0	β_1	β_2	β_3	n/D_{50}	k
Logit model						
LKB	-12.1	0.830	93.0	-57.8	2.75	
Mean dose	-13.6	0.581	90.0	-55.6		
Maximum dose	-2.71	-0.0140	66.2	-38.6		
Critical volume (4p)	-30.2	54.8	159	-99.1	12.3	14.0
Critical volume (3p)	-24.3	44.4	134	-83.3	15.4	
V_{Dth}	-22.5	41.9	148	-91.2	14.1	
Probit model						
LKB	14.0	0.156	49.7	-31.7	2.69	
Mean dose	-7.28	0.325	48.2	-30.5		
Maximum dose	-1.38	-0.00869	35.9	-21.0		
Critical volume (4p)	0.518	0.0697	76.1	-49.3	12.5	12.0
Critical volume (3p)	0.534	0.0682	77.3	-49.9	14.9	
V_{Dth}	0.501	0.0903	72.9	-46.9	14.0	

Table 7.8: Maximum likelihood estimates of the parameter values for fits of NTCP models including confounding factors to pseudo-clinical rectum dataset 5.

	β_0	β_1	β_2	β_3	n/D_{50}	k
Logit model						
LKB	-54.6	0.855	690	-642	0.315	
Mean dose	-30.5	0.586	535	-516		
Maximum dose	-33.0	0.373	330	-338		
Critical volume (4p)	-28.0	41.1	680	-649	36.2	12.7
Critical volume (3p)	-27.7	45.4	655	-593	64.4	
V_{Dth}	-24.9	34.5	643	-633	35	
Probit model						
LKB	63.0	0.0346	354	-327	0.324	
Mean dose	-16.9	0.322	286	-267		
Maximum dose	-18.1	0.203	177	-177		
Critical volume (4p)	0.699	0.0669	341	-310	36.3	12.4
Critical volume (3p)	0.751	0.0535	351	-318	36.9	
V_{Dth}	0.749	0.0760	318	-295	35.0	

Chapter 8

Conclusions

8.1 The 3D model of normal tissue response

This thesis has introduced a novel method of exploring the interface between theoretical radiobiology and clinical radiotherapy; a mechanistic 3D computer model of normal-tissue response to irradiation was developed, which simulates pseudo-clinical outcome data for arbitrary dose distributions.

The model is unavoidably simplistic, with local tissue damage represented by FSU inactivation, and the volume effect (determining organ injury) represented by a critical functioning volume (CFV). Whilst the local dose effect was assumed to be a result of cell kill, and given by the LQ model and binomial statistics, the specific mechanisms of loss of organ function, resulting from a 3D distribution of inactivated FSUs, were not specified. Instead, the very general concept of the CFV was allowed to represent all factors influencing the volume effect. In chapter 6 it was found that these mechanisms are different for different organs, which determined the size of the CFV and the organ-specific threshold of FSU inactivation within the CFV for causing a complication.

Chapter 6 was devoted to tailoring the parameter values of the model to as closely as possible represent lung and rectum as organs at risk, based on data from the literature. It was concluded that the alveolus of the lung has the appropriate number of epithelial stem cells to function as the FSU for the endpoint of radiation pneumonitis, and that the capillary cluster around a mucosal gland in the rectum is plausible as the FSU for the endpoint of late rectal bleeding. As expected, the lung was found to have a large volume effect, which to some extent depends on inflammatory processes triggered by

radiation-induced tissue damage. In contrast, there is no consensus about the volume effect of the rectum, but an intermediate size of the CFV seemed appropriate.

When the radiobiological knowledgebase of normal-tissue damage had been investigated, it was concluded that more quantitative data on local tissue damage, over a large dose range, would greatly enhance the potential for mechanistic modelling of normal-tissue complications. Determining the local tissue dose-response separately would provide a stepping stone from which the mechanisms of organ function loss, as a result of 3D distributions of FSU inactivation, could be studied. However, for many organs and endpoints the most relevant measure of local tissue function has yet to be determined.

Unlike DVH-based NTCP models, the developed 3D model can include spatial effects of dose distributions, and rather than to evaluate treatment plans its purpose is to simulate radiotherapy treatments, generating pseudo-clinical complication data for a patient depending on the choice of a set of parameters. Such simulations are ideal for studying data-analysis methodology, given the flexibility of the model and that simulations are easily performed in large numbers. However, because of the relative complexity of the model, simulations were computationally expensive, and the studies in this thesis were only feasible by running the simulations on a cluster of computers (see section 4.3).

8.2 Conclusions from studies on simulated data

The 3D model developed in this thesis was described in chapter 4, and in chapter 5 the model was used to generate pseudo-clinical data for a correlation analysis. Here three different generic organ types (parallel, semi-parallel and serial) were simulated by choosing different values for the CFV. Once the data had been generated, a correlation analysis between outcome and many different dosimetric parameters was carried out, and the results were linked to characteristics of the sets of dose distributions in the analysis.

It was found that the response of the parallel and serial organs was strongly dependent on the mean and maximum organ dose respectively, whilst no parameter related strongly to the outcome for the semi-parallel organ. In the former case all V_x and D_x parameters

correlating strongly with the outcome could be explained by a cross-correlation with the mean or maximum dose respectively. The inter-patient variability in dosimetric parameters was also shown to influence the strength of correlation observed. However, it was found that the volume, x, for which the D_x parameter correlated the strongest with outcome for the serial and semi-parallel organs, corresponded to the size of the CFV. This indicates that some information of the volume effect of an organ can be gained from a correlation analysis, if care is taken to identify other factors influencing the observed correlations.

After having derived plausible values for the 3D model parameters in chapter 6, in order to represent two real organs, in chapter 7 simulations were carried out for radiation pneumonitis and late rectal bleeding as endpoints, with dose distributions representative of lung and prostate cancer treatments. Various DVH-based NTCP models were then fitted to the generated pseudo-clinical datasets. These datasets included different levels of confounding factors (inter-patient variation in both radiosensitivity and health status), which allowed the performance of the models to be compared under various conditions. The results showed that the influence of confounding factors on the performance of DVH-based models is more important than the choice of model. Further, methodologies for accounting for non-dosimetric factors were described, and the importance of finding ways of quantifying such factors was emphasised.

8.3 Future work

Since the studies in chapters 5 and 7 were intended to explore effects influencing analyses of clinical data, the datasets generated for these studies were made to resemble clinical datasets, in terms of inter-patient variation in doses and volumes treated. Typically, each patient had a unique, automatically generated, dose distribution, and in some datasets each patient also had a unique value of α and N_0^{FSU} (sampled from a lognormal / log-lognormal distribution respectively), representing radiosensitivity and health status.

Consequently, the data simulated and analysed in this thesis were characterised largely by the dose distributions. However, due to its structure the model lends itself to explore several other phenomena important to radiotherapy; future applications of the model could include initially inhomogeneous FSU distributions, representing e.g.

the spatial distribution of organ damage detected by functional imaging, as well as modifications to the cell kill function to include low-dose hypersensitivity and time-effects. It would also be of special interest to introduce non-local effects of irradiation into the model (i.e. radiation-induced bystander effects; see e.g. [196]), such that the cell kill is determined not only by the local dose, but also by the dose to cells in the vicinity.

Just as the performance of DVH-based NTCP models was tested on data simulated by the model, so that the fitted model parameters could be compared to the 'actual' volume effect and local dose-response function (chapter 7), also the feasibility of retrieving the α/β-ratio of the LQ-model from 'clinical' datasets of different fractionation could be tested. Focusing further on fractionation, it would also be of interest to generate Strandqvist isotoxicity curves for different sizes of the CFV, and test different models' ability to predict the results.

The 3D model will further be adapted to be able to simulate treatments based on actual treatment plans. The full dose cubes and structure information for prostate treatments could be imported into the model, and rather than defining the rectum as a square tube, as in chapter 6, the outline of the real rectum would be used. In this way the plausibility of the model parameters derived for rectum could be tested further. Also, different endpoints relevant for one and the same organ, like late rectal bleeding and loose stools, can be modelled simultaneously by applying the same dose distribution to two different FSU distributions which co-exist in the same voxels. The size of the CFV is expected to be different for these different endpoints, as well as the dose response of the FSU populations.

It might be convenient to adapt the model code to be able to run simulations in parallel on a GPU, rather than on a dedicated cluster of computers. This would potentially greatly simplify the logistics of using the model.

Bibliography

[1] Hendry, J. H., Jeremić, B. and Zubizarreta, E. H. Normal tissue complications after radiation therapy. *Rev Panam Salud Publica* **20**, 151–160 (2006).

[2] Gagliardi, G. *et al.* Radiation dose-volume effects in the heart. *Int J Radiat Oncol Biol Phys* **76**, S77–S85 (2010).

[3] Kirkpatrick, J. P., van der Kogel, A. J. and Schultheiss, T. E. Radiation dose-volume effects in the spinal cord. *Int J Radiat Oncol Biol Phys* **76**, S42–S49 (2010).

[4] Sarna, L. *et al.* Impact of respiratory symptoms and pulmonary function on quality of life of long-term survivors of non-small cell lung cancer. *Chest* **125**, 439–445 (2004).

[5] Dunberger, G. *et al.* Fecal incontinence affecting quality of life and social functioning among long-term gynecological cancer survivors. *Int J Gynecol Cancer* **20**, 449–460 (2010).

[6] Deasy, J. O. *et al.* Radiotherapy dose-volume effects on salivary gland function. *Int J Radiat Oncol Biol Phys* **76**, S58–S63 (2010).

[7] Rancati, T. *et al.* Radiation dose-volume effects in the larynx and pharynx. *Int J Radiat Oncol Biol Phys* **76**, S64–S69 (2010).

[8] Werner-Wasik, M. *et al.* Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys* **76**, S86–S93 (2010).

[9] Nias, A. H. W. *An introduction to Radiobiology*, 2 edn. (Wiley, Chichester, 1990).

[10] Phillips, T. L. Radiation fibrosis, *Pulmonary Diseases and Disorders*, 2 edn. Fishman, A. P. (ed.). chap. 51, 773–792 (McGraw-Hill, New York, 1988).

[11] Steel, G. Radiobiology of tumours, *Handbook of Radiotherapy Physics.* Mayles, P., Nahum, A. E. and Rosenwald, J. (eds.). chap. 7, 127–148 (Taylor & Francis, Boca Raton ; London, 2007).

[12] Hendry, J. H. Response of human organs to single (or fractionated equivalent) doses of irradiation. *Int J Radiat Biol* **56**, 691–700 (1989).

[13] Rubin, P. The Franz Buschke lecture: late effects of chemotherapy and radiation therapy: a new hypothesis. *Int J Radiat Oncol Biol Phys* **10**, 5–34 (1984).

[14] Partridge, M. A radiation damage repair model for normal tissues. *Phys Med Biol* **53**, 3595–3608 (2008).

[15] Steel, G. Dose fractionation in radiotherapy, *Handbook of Radiotherapy Physics.* Mayles, P., Nahum, A. E. and Rosenwald, J. (eds.). chap. 9, 163–178 (Taylor & Francis, Boca Raton ; London, 2007).

[16] Brenner, D. J. The linear-quadratic model is an appropriate methodology for determining isoeffective doses at large doses per fraction. *Semin Radiat Oncol* **18**, 234–239 (2008).

[17] Thames, H. D. *et al.* Fractionation parameters for human tissues and tumors. *Int J Radiat Biol* **56**, 701–710 (1989).

[18] Thames, H. D. and Hendry, J. H. *Fractionation in Radiotherapy* (Taylor and Francis, London and Philadelphia, 1987).

[19] Stewart, F. A. and Dörr, W. Milestones in normal tissue radiation biology over the past 50 years: from clonogenic cell survival to cytokine networks and back to stem cell recovery. *Int J Radiat Biol* **85**, 574–586 (2009).

[20] Hill, R. P. *et al.* Normal tissue radiobiology: from the laboratory to the clinic. *Int J Radiat Oncol Biol Phys* **49**, 353–365 (2001).

[21] Travis, E. L. Organizational response of normal tissues to irradiation. *Semin Radiat Oncol* **11**, 184–196 (2001).

[22] Coleman, C. N. International Conference on Translational Research and Preclinical Strategies in Radio-Oncology (ICTR)–conference summary. *Int J Radiat Oncol Biol Phys* **49**, 301–309 (2001).

[23] Rubin, E. and Farber, J. *Pathology*, 3 edn. (Lippincott-Raven, Philadelphia, 1999).

[24] Fajardo, L. F. The pathology of ionizing radiation as defined by morphologic patterns. *Acta Oncol* **44**, 13–22 (2005).

[25] Denham, J. W. and Hauer-Jensen, M. The radiotherapeutic injury–a complex 'wound'. *Radiother Oncol* **63**, 129–145 (2002).

[26] Hopewell, J. W. and Trott, K. R. Volume effects in radiobiology as applied to radiotherapy. *Radiother Oncol* **56**, 283–288 (2000).

[27] Marks, L. B. The impact of organ structure on radiation response. *Int J Radiat Oncol Biol Phys* **34**, 1165–1171 (1996).

[28] Steel, G. Radiobiology of normal tissues, *Handbook of Radiotherapy Physics.* Mayles, P., Nahum, A. E. and Rosenwald, J. (eds.). chap. 8, 149–162 (Taylor & Francis, Boca Raton ; London, 2007).

[29] Withers, H. R., Taylor, J. M. and Maciejewski, B. Treatment volume and tissue tolerance. *Int J Radiat Oncol Biol Phys* **14**, 751–759 (1988).

[30] Alber, M. and Nüsslin, F. An objective function for radiation treatment optimization based on local biological measures. *Phys Med Biol* **44**, 479–493 (1999).

[31] Fenwick, J. D. and Nahum, A. E. Series model volume effects in a population of non-identical patients: how low is low? *Phys Med Biol* **46**, 1815–1834 (2001).

[32] Raphael, C. Mathematical modelling of objectives in radiation therapy treatment planning. *Phys Med Biol* **37**, 1293–1311 (1992).

[33] Schultheiss, T. E. The controversies and pitfalls in modeling normal tissue radiation injury/damage. *Semin Radiat Oncol* **11**, 210–214 (2001).

[34] Niemierko, A. and Goitein, M. Modeling of normal tissue response to radiation: the critical volume model. *Int J Radiat Oncol Biol Phys* **25**, 135–145 (1993).

[35] Yorke, E. D. *et al.* Probability of radiation-induced complications in normal tissues with parallel architecture under conditions of uniform whole or partial organ irradiation. *Radiother Oncol* **26**, 226–237 (1993).

[36] Timmerman, R. *et al.* Optimizing dose and fractionation for stereotactic body radiation therapy. Normal tissue and tumor control effects with large dose per fraction. *Front Radiat Ther Oncol* **40**, 352–365 (2007).

[37] Milano, M. T., Constine, L. S. and Okunieff, P. Normal tissue toxicity after small field hypofractionated stereotactic body radiation. *Radiat Oncol* **3**, 36 (2008).

[38] Paris, F. *et al.* Endothelial apoptosis as the primary lesion initiating intestinal radiation damage in mice. *Science* **293**, 293–297 (2001).

[39] Novakova-Jiresova, A. *et al.* Pulmonary radiation injury: identification of risk factors associated with regional hypersensitivity. *Cancer Res* **65**, 3568–3576 (2005).

[40] Philippens, M. E. P. *et al.* Dose-volume effects in rat thoracolumbar spinal cord: an evaluation of NTCP models. *Int J Radiat Oncol Biol Phys* **60**, 578–590 (2004).

[41] Tucker, S. L., Liao, Z. X. and Travis, E. L. Estimation of the spatial distribution of target cells for radiation pneumonitis in mouse lung. *Int J Radiat Oncol Biol Phys* **38**, 1055–1066 (1997).

[42] Alber, M. Normal tissue dose-effect models in biological dose optimisation. *Z Med Phys* **18**, 102–110 (2008).

[43] Seppenwoolde, Y. *et al.* Comparing different NTCP models that predict the incidence of radiation pneumonitis. Normal tissue complication probability. *Int J Radiat Oncol Biol Phys* **55**, 724–735 (2003).

[44] Tucker, S. L. *et al.* Dose-volume response analyses of late rectal bleeding after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* **59**, 353–365 (2004).

[45] Collett, D. *Modelling binary data*, 2 edn. (Chapman & Hall/CRC, 2003).

[46] Deasy, J. O. *et al.* Methodological issues in radiation dose-volume outcome analyses: summary of a joint AAPM/NIH workshop. *Med Phys* **29**, 2109–2127 (2002).

[47] Källman, P., Ågren, A. and Brahme, A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int J Radiat Biol* **62**, 249–262 (1992).

[48] Bentzen, S. M. and Tucker, S. L. Quantifying the position and steepness of radiation dose-response curves. *Int J Radiat Biol* **71**, 531–542 (1997).

[49] Rancati, T. *et al.* Fitting late rectal bleeding data using different NTCP models: results from an Italian multi-centric study (AIROPROS0101). *Radiother Oncol* **73**, 21–32 (2004).

[50] Kutcher, G. J. and Burman, C. Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method. *Int J Radiat Oncol Biol Phys* **16**, 1623–1630 (1989).

[51] Lyman, J. T. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl* **8**, S13–S19 (1985).

[52] Stavrev, P. *et al.* Critical volume model analysis of lung complication data from different strains of mice. *Int J Radiat Biol* **81**, 77–88 (2005).

[53] Tsougos, I. *et al.* Evaluation of dose-response models and parameters predicting radiation induced pneumonitis using clinical data from breast cancer radiotherapy. *Phys Med Biol* **50**, 3535–3554 (2005).

[54] Niemierko, A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. *Med Phys* **24**, 103–110 (1997).

[55] Mohan, R. *et al.* Clinically relevant optimization of 3-D conformal treatments. *Med Phys* **19**, 933–944 (1992).

[56] Niemierko, A. A generalized concept of equivalent uniform dose (EUD). *Med Phys* **26**, 1100 (1999).

[57] Mavroidis, P., Lind, B. K. and Brahme, A. Biologically effective uniform dose ($\bar{\bar{D}}$) for specification, report and comparison of dose response relations and treatment plans. *Phys Med Biol* **46**, 2607–2630 (2001).

[58] Schultheiss, T. E., Orton, C. G. and Peck, R. A. Models in radiotherapy: volume effects. *Med Phys* **10**, 410–415 (1983).

[59] Boersma, L. J. *et al.* Dose-effect relations for local functional and structural changes of the lung after irradiation for malignant lymphoma. *Radiother Oncol* **32**, 201–209 (1994).

[60] van Luijk, P. *et al.* Estimation of parameters of dose-volume models and their confidence limits. *Phys Med Biol* **48**, 1863–1884 (2003).

[61] Jin, J.-Y. *et al.* Impact of fraction size on lung radiation toxicity: hypofractionation may be beneficial in dose escalation of radiotherapy for lung cancers. *Int J Radiat Oncol Biol Phys* **76**, 782–788 (2010).

[62] Jackson, A., Kutcher, G. J. and Yorke, E. D. Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation. *Med Phys* **20**, 613–625 (1993).

[63] van Luijk, P. *et al.* Bath and shower effects in the rat parotid gland explain increased relative risk of parotid gland dysfunction after intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys* **74**, 1002–1005 (2009).

[64] Stavrev, P. *et al.* Generalization of a model of tissue response to radiation based on the idea of functional subunits and binomial statistics. *Phys Med Biol* **46**, 1501–1518 (2001).

[65] Stavreva, N. *et al.* Modelling the dose-volume response of the spinal cord, based on the idea of damage to contiguous functional subunits. *Int J Radiat Biol* **77**, 695–702 (2001).

[66] van Luijk, P. *et al.* Data on dose-volume effects in the rat spinal cord do not support existing NTCP models. *Int J Radiat Oncol Biol Phys* **61**, 892–900 (2005).

[67] Thames, H. D. *et al.* Cluster models of dose-volume effects. *Int J Radiat Oncol Biol Phys* **59**, 1491–1504 (2004).

[68] Tucker, S. L. *et al.* Cluster model analysis of late rectal bleeding after IMRT of prostate cancer: a case-control study. *Int J Radiat Oncol Biol Phys* **64**, 1255–1264 (2006).

[69] Fenwick, J. Ph.D. thesis, University of London (1999).

[70] Roach, M. *et al.* The "critical volume tolerance method" for estimating the limits of dose escalation during three-dimensional conformal radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* **35**, 1019–1025 (1996).

[71] Bonta, D. V. *et al.* A variable critical-volume model for normal tissue complication probability. *Med Phys* **28**, 1338–1343 (2001).

[72] Withers, H. R. and Taylor, J. M. Critical volume model. *Int J Radiat Oncol Biol Phys* **25**, 151–152 (1993).

[73] Deasy, J. O. *et al.* Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. *Int J Radiat Oncol Biol Phys* **76**, S151–S154 (2010).

[74] Marks, L. B. *et al.* Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys* **76**, S10–S19 (2010).

[75] Tucker, S. L. *et al.* Dose-volume modeling of the risk of postoperative pulmonary complications among esophageal cancer patients treated with concurrent chemoradiotherapy followed by surgery. *Int J Radiat Oncol Biol Phys* **66**, 754–761 (2006).

[76] Pettersson, N., Nyman, J. and Johansson, K.-A. Radiation-induced rib fractures after hypofractionated stereotactic body radiation therapy of non-small cell lung cancer: a dose- and volume-response analysis. *Radiother Oncol* **91**, 360–368 (2009).

[77] Fowler, J. F., Welsh, J. S. and Howard, S. P. Loss of biological effect in prolonged fraction delivery. *Int J Radiat Oncol Biol Phys* **59**, 242–249 (2004).

[78] Dale, R. G. The application of the linear-quadratic dose-effect equation to fractionated and protracted radiotherapy. *Br J Radiol* **58**, 515–528 (1985).

[79] Goitein, M. How best to dispose of extra-tumoral dose: a cautionary note for intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* **75**, 1–3 (2009).

[80] Zaider, M. and Amols, H. I. A little to a lot or a lot to a little: is NTCP always minimized in multiport therapy? *Int J Radiat Oncol Biol Phys* **41**, 945–950 (1998).

[81] Bentzen, S. M. *et al.* Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys* **76**, S3–S9 (2010).

[82] Willner, J. *et al.* A little to a lot or a lot to a little? An analysis of pneumonitis risk from dose-volume histogram parameters of the lung in patients with lung cancer treated with 3-D conformal radiotherapy. *Strahlenther Onkol* **179**, 548–556 (2003).

[83] Rice, J. A. *Mathematical statistics and data analysis*, 2 edn. (Duxbury Press, 1995).

[84] Gagliardi, G. Ph.D. thesis, Stockholm University, Stockholm (1998).

[85] Schilstra, C. and Meertens, H. Calculation of the uncertainty in complication probability for various dose-response models, applied to the parotid gland. *Int J Radiat Oncol Biol Phys* **50**, 147–158 (2001).

[86] Rutkowska, E. *et al.* Confidence intervals on NTCP estimates for individualised dose-prescription. *Radiat Oncol* **88**, S29 (2008). ESTRO conference poster (Gothenburg).

[87] Nahum, A. E. and Sanchez-Nieto, B. Tumour control probability modelling: Basic principles and applications in treatment planning. *Physica Medica* **17**, 13–23 (2001).

[88] Malik, Z. *et al.* Iso-NTCP customisation of the prescription dose in lung-tumour radiotherapy. *Radiother Oncol* **84**, S278–S279 (2007). ESTRO conference poster (Barcelona).

[89] Graham, M. V. *et al.* Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). *Int J Radiat Oncol Biol Phys* **45**, 323–329 (1999).

[90] Kong, F.-M. *et al.* Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): predictors for radiation pneumonitis and fibrosis. *Int J Radiat Oncol Biol Phys* **65**, 1075–1086 (2006).

[91] Kwa, S. L. *et al.* Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients. *Int J Radiat Oncol Biol Phys* **42**, 1–9 (1998).

[92] Rancati, T. *et al.* Factors predicting radiation pneumonitis in lung cancer patients: a retrospective study. *Radiother Oncol* **67**, 275–283 (2003).

[93] De Jaeger, K. *et al.* Incorporating an improved dose-calculation algorithm in conformal radiotherapy of lung cancer: re-evaluation of dose in normal lung tissue. *Radiother Oncol* **69**, 1–10 (2003).

[94] Lind, P. A. *et al.* ROC curves and evaluation of radiation-induced pulmonary toxicity in breast cancer. *Int J Radiat Oncol Biol Phys* **64**, 765–770 (2006).

[95] Zhou, X.-H., Obuchowski, N. A. and McClish, D. K. *Statistical methods in diagnostic medicine* (John Wiley & Sons, New York, 2002).

[96] Jackson, A. *et al.* The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome. *Int J Radiat Oncol Biol Phys* **76**, S155–S160 (2010).

[97] DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

[98] Akobeng, A. K. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* **96**, 644–647 (2007).

[99] Burnham, K. P. and Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*, 2 edn. (Springer-Verlag, New York, 2002).

[100] El Naqa, I. *et al.* Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors. *Int J Radiat Oncol Biol Phys* **64**, 1275–1286 (2006).

[101] Houweling, A. C. *et al.* A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* **76**, 1259–1265 (2010).

[102] Tucker, S. L. *et al.* Comparison of rectal dose-wall histogram versus dose-volume histogram for modeling the incidence of late rectal bleeding after radiotherapy. *Int J Radiat Oncol Biol Phys* **60**, 1589–1601 (2004).

[103] Efron, B. and Tibshirani, R. J. *An introduction to the bootstrap* (Chapman & Hall/CRC, 1998).

[104] Jackson, A. *et al.* Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* **31**, 883–891 (1995).

[105] El Naqa, I. *et al.* Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. *Phys Med Biol* **51**, 5719–5735 (2006).

[106] Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).

[107] Rutkowska, E., Baker, C. and Nahum, A. Mechanistic simulation of normal-tissue damage in radiotherapy–implications for dose-volume analyses. *Phys Med Biol* **55**, 2121–2136 (2010).

[108] Nahum, A. E. and Kutcher, G. J. Biological evaluation of treatment plans, *Handbook of Radiotherapy Physics.* Mayles, P., Nahum, A. E. and Rosenwald, J. (eds.). chap. 36, 731–772 (Taylor & Francis, Boca Raton ; London, 2007).

[109] Johnson, T. D. *et al.* A Bayesian mixture model relating dose to critical organs and functional complication in 3D conformal radiation therapy. *Biostatistics* **6**, 615–632 (2005).

[110] Jordan, T. J. Megavoltage X-ray beams: 2-50 MV. *BJR Suppl* **25**, 62–109 (1996).

[111] Wolbarst, A. B., Chin, L. M. and Svensson, G. K. Optimization of radiation therapy: integral-response of a model biological system. *Int J Radiat Oncol Biol Phys* **8**, 1761–1769 (1982).

[112] Wolbarst, A. B. Optimization of radiation therapy II: the critical-voxel model. *Int J Radiat Oncol Biol Phys* **10**, 741–745 (1984).

[113] Buettner, F. *et al.* Assessing correlations between the spatial distribution of the dose to the rectal wall and late rectal toxicity after prostate radiotherapy: an analysis of data from the MRC RT01 trial (ISRCTN 47772397). *Phys Med Biol* **54**, 6535–6548 (2009).

[114] Alber, M. and Belka, C. A normal tissue dose response model of dynamic repair processes. *Phys Med Biol* **51**, 153–172 (2006).

[115] Fenwick, J. D. An ntcp formula for the series model with a Yaes-type functional subunit mechanism. *Phys Med Biol* **46**, N33–N38 (2001).

[116] Chapman, J. D. Single-hit mechanism of tumour cell killing by radiation. *Int J Radiat Biol* **79**, 71–81 (2003).

[117] Mothersill, C., Seymour, C. B. and Joiner, M. C. Relationship between radiation-induced low-dose hypersensitivity and the bystander effect. *Radiat Res* **157**, 526–532 (2002).

[118] Kong, F.-M. S. *et al.* Physical models and simpler dosimetric descriptors of radiation late toxicity. *Semin Radiat Oncol* **17**, 108–120 (2007).

[119] Hendry, J. H. and Moore, J. V. Deriving absolute values of alpha and beta for dose fractionation, using dose-incidence data. *Br J Radiol* **58**, 885–890 (1985).

[120] Bentzen, S. M. and Overgaard, M. Relationship between early and late normal-tissue injury after postmastectomy radiotherapy. *Radiother Oncol* **20**, 159–165 (1991).

[121] Kehwar, T. S. Analytical approach to estimate normal tissue complication probability using best fit of normal tissue tolerance doses into the NTCP equation of the linear quadratic model. *J Cancer Res Ther* **1**, 168–179 (2005).

[122] Weibel, E. R. and Taylor, C. R. Design and structure of the human lung, *Pulmonary Diseases and Disorders*, 2 edn. Fishman, A. P. (ed.). vol. 1, chap. 2 (McGraw-Hill, New York, 1988).

[123] Milano, M. T., Constine, L. S. and Okunieff, P. Normal tissue tolerance dose metrics for radiation therapy of major organs. *Semin Radiat Oncol* **17**, 131–140 (2007).

[124] Joiner, M. Quantifying cell kill and cell survival, *Basic clinical radiobiology*, 4th edn. Joiner, M. and van der Kogel, A. (eds.). chap. 4, 41–55 (Hodder Arnold, 2009).

[125] Dörr, W. Pathogenesis of normal-tissue side-effects, *Basic clinical radiobiology*, 4th edn. Joiner, M. and van der Kogel, A. (eds.). chap. 13, 169–190 (Hodder Arnold, 2009).

[126] Travis, E. L. and Terry, N. H. Cell depletion and initial and chronic responses in normal tissues. *Front Radiat Ther Oncol* **23**, 41–59 (1989).

[127] Zhao, W., Diz, D. I. and Robbins, M. E. Oxidative damage pathways in relation to normal tissue injury. *Br J Radiol* **80 Spec No 1**, S23–S31 (2007).

[128] Barber, J. and James, R. Radiation effects on the human gastrointestinal tract, *Radiation and gut*. Potten, C. and Hendry, J. H. (eds.). chap. 8, 231–252 (Elsevier Science B.V., The Netherlands, 1995).

[129] Wang, J. *et al.* Deficiency of microvascular thrombomodulin and up-regulation of protease-activated receptor-1 in irradiated rat intestine: possible link between endothelial dysfunction and chronic radiation fibrosis. *Am J Pathol* **160**, 2063–2072 (2002).

[130] Marks, L. B. *et al.* Radiation dose-volume effects in the lung. *Int J Radiat Oncol Biol Phys* **76**, S70–S76 (2010).

[131] Ferlay, J. *et al.* Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol* **18**, 581–592 (2007).

[132] Tsoutsou, P. G. and Koukourakis, M. I. Radiation pneumonitis and fibrosis: mechanisms underlying its pathogenesis and implications for future research. *Int J Radiat Oncol Biol Phys* **66**, 1281–1293 (2006).

[133] Gopal, R. *et al.* The relationship between local dose and loss of function for irradiated lung. *Int J Radiat Oncol Biol Phys* **56**, 106–113 (2003).

[134] Fan, M. *et al.* Can we predict radiation-induced changes in pulmonary function based on the sum of predicted regional dysfunction? *J Clin Oncol* **19**, 543–550 (2001).

[135] Seppenwoolde, Y. *et al.* Radiation dose-effect relations and local recovery in perfusion for patients with non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* **47**, 681–690 (2000).

[136] Theuws, J. C. *et al.* Dose-effect relations for early local pulmonary injury after irradiation for malignant lymphoma and breast cancer. *Radiother Oncol* **48**, 33–43 (1998).

[137] Marks, L. B. and Ma, J. Challenges in the clinical application of advanced technologies to reduce radiation-associated normal tissue injury. *Int J Radiat Oncol Biol Phys* **69**, 4–12 (2007).

[138] Allen, A. M. *et al.* Fatal pneumonitis associated with intensity-modulated radiation therapy for mesothelioma. *Int J Radiat Oncol Biol Phys* **65**, 640–645 (2006).

[139] Wang, S.-l. *et al.* Investigation of clinical and dosimetric factors associated with postoperative pulmonary complications in esophageal cancer patients treated with concurrent chemoradiotherapy followed by surgery. *Int J Radiat Oncol Biol Phys* **64**, 692–699 (2006).

[140] Morgan, T. L. *et al.* A comparison of single-dose and fractionated total-body irradiation on the development of pneumonitis following bone marrow transplantation. *Int J Radiat Oncol Biol Phys* **36**, 61–66 (1996).

[141] Thames, H. D. *et al.* The high steepness of dose-response curves for late-responding normal tissues. *Radiother Oncol* **15**, 49–53 (1989).

[142] Lumb, A. B. and Nunn, J. F. *Nunn's applied respiratory physiology*, 6 edn. (Elsevier Butterworth Heinemann, Philadelphia, Pa., 2005).

[143] Stone, K. C. *et al.* Allometric relationships of cell numbers and size in the mammalian lung. *Am J Respir Cell Mol Biol* **6**, 235–243 (1992).

[144] Fehrenbach, H. Alveolar epithelial type II cell: defender of the alveolus revisited. *Respir Res* **2**, 33–46 (2001).

[145] Ochs, M. *et al.* The number of alveoli in the human lung. *Am J Respir Crit Care Med* **169**, 120–124 (2004).

[146] Hart, J. P. *et al.* Radiation pneumonitis: correlation of toxicity with pulmonary metabolic radiation response. *Int J Radiat Oncol Biol Phys* **71**, 967–971 (2008).

[147] Asakura, H. *et al.* Analysis of dose-volume histogram parameters for radiation pneumonitis after definitive concurrent chemoradiotherapy for esophageal cancer. *Radiother Oncol* **95**, 240–244 (2010).

[148] Dehing-Oberije, C. *et al.* The importance of patient characteristics for the prediction of radiation-induced lung toxicity. *Radiother Oncol* **91**, 421–426 (2009).

[149] Moon, S. H. *et al.* Radiation-induced pulmonary toxicity following adjuvant radiotherapy for breast cancer. *J Korean Soc Ther Radiol Oncol* **25**, 109–117 (2007).

[150] Oh, D. *et al.* Prediction of radiation pneumonitis following high-dose thoracic radiation therapy by 3 Gy/fraction for non-small cell lung cancer: analysis of clinical and dosimetric factors. *Jpn J Clin Oncol* **39**, 151–157 (2009).

[151] Schallenkamp, J. M. *et al.* Incidence of radiation pneumonitis after thoracic irradiation: Dose-volume correlates. *Int J Radiat Oncol Biol Phys* **67**, 410–416 (2007).

[152] Hernando, M. L. *et al.* Radiation-induced pulmonary toxicity: a dose-volume histogram analysis in 201 patients with lung cancer. *Int J Radiat Oncol Biol Phys* **51**, 650–659 (2001).

[153] Yorke, E. D. *et al.* Dose-volume factors contributing to the incidence of radiation pneumonitis in non-small-cell lung cancer patients treated with three-dimensional conformal radiation therapy. *Int J Radiat Oncol Biol Phys* **54**, 329–339 (2002).

[154] Becciolini, A. Relative radiosensitivities of the small and large intestine, *Relative radiation sensitivities of human organ systems.* Lett, J. T. and Altman, K. I. (eds.). vol. 12 of *Advances in radiation biology*, chap. 2, 83–128 (Academic Press, New York, 1987).

[155] Breen, E. and Brooks, D. C. Management of radiation injury to the bowel, *Prostate cancer: principles and practice.* Kantoff, P. W., Carroll, P. R. and D'Amico, A. V. (eds.). chap. 26B, 392–396 (Lippincott Williams & Wilkins, 2002).

[156] Colwell, J. C. and Goldberg, M. A review of radiation proctitis in the treatment of prostate cancer. *J Wound Ostomy Continence Nurs* **27**, 179–187 (2000).

[157] Michalski, J. M. *et al.* Radiation dose-volume effects in radiation-induced rectal injury. *Int J Radiat Oncol Biol Phys* **76**, S123–S129 (2010).

[158] O'Brien, P. C. Radiation injury of the rectum. *Radiother Oncol* **60**, 1–14 (2001).

[159] Phan, J. *et al.* Late rectal complications after prostate brachytherapy for localized prostate cancer: incidence and management. *Cancer* **115**, 1827–1839 (2009).

[160] Zelefsky, M. J. *et al.* Long term tolerance of high dose three-dimensional conformal radiotherapy in patients with localized prostate carcinoma. *Cancer* **85**, 2460–2468 (1999).

[161] van Lin, E. N. *et al.* Reduced late rectal mucosal changes after prostate three-dimensional conformal radiotherapy with endorectal balloon as observed in repeated endoscopy. *Int J Radiat Oncol Biol Phys* **67**, 799–811 (2007).

[162] Kruse, J. J. C. M. *et al.* Microarray analysis to identify molecular mechanisms of radiation-induced microvascular damage in normal tissues. *Int J Radiat Oncol Biol Phys* **58**, 420–426 (2004).

[163] Reinhold, H. S. The influence of radiation on blood vessels and circulation. Chapter IV. Structural changes in blood vessels. *Curr Top Radiat Res Q* **10**, 58–74 (1974).

[164] Fajardo, L. F. and Stewart, J. R. Pathogenesis of radiation-induced myocardial fibrosis. *Lab Invest* **29**, 244–257 (1973).

[165] Haboubi, N. Y., Schofield, P. F. and Rowland, P. L. The light and electron microscopic features of early and late phase radiation-induced proctitis. *Am J Gastroenterol* **83**, 1140–1144 (1988).

[166] Stillie, A. L. *et al.* Rectal filling at planning does not predict stability of the prostate gland during a course of radical radiotherapy if patients with large rectal filling are re-imaged. *Clin Oncol (R Coll Radiol)* **21**, 760–767 (2009).

[167] Rasmussen, S. N. and Riis, P. Rectal wall thickness measured by ultrasound in chronic inflammatory diseases of the colon. *Scand J Gastroenterol* **20**, 109–114 (1985).

[168] Fait, E. *et al.* Microvascular patterns of the human large intestine: morphometric studies of vascular parameters in corrosion casts. *Scanning Microscopy* **12**, 641–651 (1998).

[169] Urbich, C. *et al.* Shear stress-induced endothelial cell migration involves integrin signaling via the fibronectin receptor subunits alpha(5) and beta(1). *Arterioscler Thromb Vasc Biol* **22**, 69–75 (2002).

[170] Kessel, R. G. *Basic medical histology* (Oxford University Press, New York, 1998).

[171] Schor, A. M., Schor, S. L. and Arciniegas, E. Phenotypic diversity and lineage relationships in vasculare endothelial cells, *Stem cells.* Potten (ed.). chap. 6 (Academic Press, 1997).

[172] Asahara, T. and Isner, J. M. Endothelial progenitor cells for vascular regeneration. *J Hematother Stem Cell Res* **11**, 171–178 (2002).

[173] Pober, J. S. and Min, W. Endothelial cell dysfunction, injury and death. *Handb Exp Pharmacol* 135–156 (2006).

[174] Gajdusek, C. *et al.* Early molecular changes in irradiated aortic endothelium. *J Cell Physiol* **188**, 8–23 (2001).

[175] Chang, E. *et al.* Aging and survival of cutaneous microvasculature. *J Invest Dermatol* **118**, 752–758 (2002).

[176] van der Kogel, A. J. *et al.* Radiation tolerance of the rat rectum to fractionated X-rays and pi-mesons. *Radiother Oncol* **12**, 225–232 (1988).

[177] Bentzen, S. M., Turesson, I. and Thames, H. D. Fractionation sensitivity and latency of telangiectasia after postmastectomy radiotherapy: a graded-response analysis. *Radiother Oncol* **18**, 95–106 (1990).

[178] Wachter, S. *et al.* Endoscopic scoring of late rectal mucosal damage after conformal radiotherapy for prostatic carcinoma. *Radiother Oncol* **54**, 11–19 (2000).

[179] Haimovitz-Friedman, A. and Fuks, Z. Signaling in the radiation response of endothelial cells, *The radiation biology of the vascular endothelium.* Rubin, D. B. (ed.). chap. 6 (CRC press, Boca Raton, 1998).

[180] Begg, A. C. and Terry, N. H. The sensitivity of normal stroma to fractionated radiotherapy measured by a tumour growth rate assay. *Radiother Oncol* **2**, 333–341 (1984).

[181] Brenner, D. J. Fractionation and late rectal toxicity. *Int J Radiat Oncol Biol Phys* **60**, 1013–1015 (2004).

[182] Dewit, L. and Oussoren, Y. Late effects in the mouse small intestine after a clinically relevant multifractionated radiation treatment. *Radiat Res* **110**, 372–384 (1987).

[183] Fiorino, C. *et al.* Dose-volume effects for normal tissues in external radiotherapy: pelvis. *Radiother Oncol* **93**, 153–167 (2009).

[184] Reinhold, H. S. Late vascular damage and its importance for radiotherapy. *Bibl Anat* 307–310 (1977).

[185] Osborne, J. Early and late radiation effects (external irradiation) on the gut, *Radiation and gut.* Potten, C. and Hendry, J. H. (eds.). chap. 6, 145–209 (Elsevier Science B.V., The Netherlands, 1995).

[186] Trott, K. R., Breiter, N. and Spiethoff, A. Experimental studies on the pathogenesis of the chronic radiation ulcer of the large bowel in rats. *Int J Radiat Oncol Biol Phys* **12**, 1637–1643 (1986).

[187] Safwat, A. *et al.* Deterministic rather than stochastic factors explain most of the variation in the expression of skin telangiectasia after radiotherapy. *Int J Radiat Oncol Biol Phys* **52**, 198–204 (2002).

[188] Valdagni, R. *et al.* To bleed or not to bleed. A prediction based on individual gene profiling combined with dose-volume histogram shapes in prostate cancer patients undergoing three-dimensional conformal radiation therapy. *Int J Radiat Oncol Biol Phys* **74**, 1431–1440 (2009).

[189] Blanco, A. I. *et al.* Dose-volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy. *Int J Radiat Oncol Biol Phys* **62**, 1055–1069 (2005).

[190] Gayou, O. *et al.* A genetic algorithm for variable selection in logistic regression analysis of radiotherapy treatment outcomes. *Med Phys* **35**, 5426–5433 (2008).

[191] Tucker, S. L. *et al.* Analysis of radiation pneumonitis risk using a generalized Lyman model. *Int J Radiat Oncol Biol Phys* **72**, 568–574 (2008).

[192] Partridge, M. and Buettner, F. Artificial neural networks for toxicity prediction. *Radiother Oncol* **96**, S107 (2010).

[193] Peeters, S. T. H. *et al.* Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling. *Int J Radiat Oncol Biol Phys* **66**, 11–19 (2006).

[194] McCullagh, P. and Nelder, J. A. *Genearalized linear models* (Chapman and Hall, London, New York, 1983).

[195] Defraene, G. *et al.* Importance of including clinical factors in rectal NTCP modeling after radiotherapy for prostate cancer. *Radiother Oncol* **96**, S127 (2010).

[196] Chai, Y. and Hei, T. K. Radiation induced bystander effect in vivo. *Acta Med Nagasaki* **53**, S65–S69 (2008).